# Data Mining II

## Mobility Data Mining

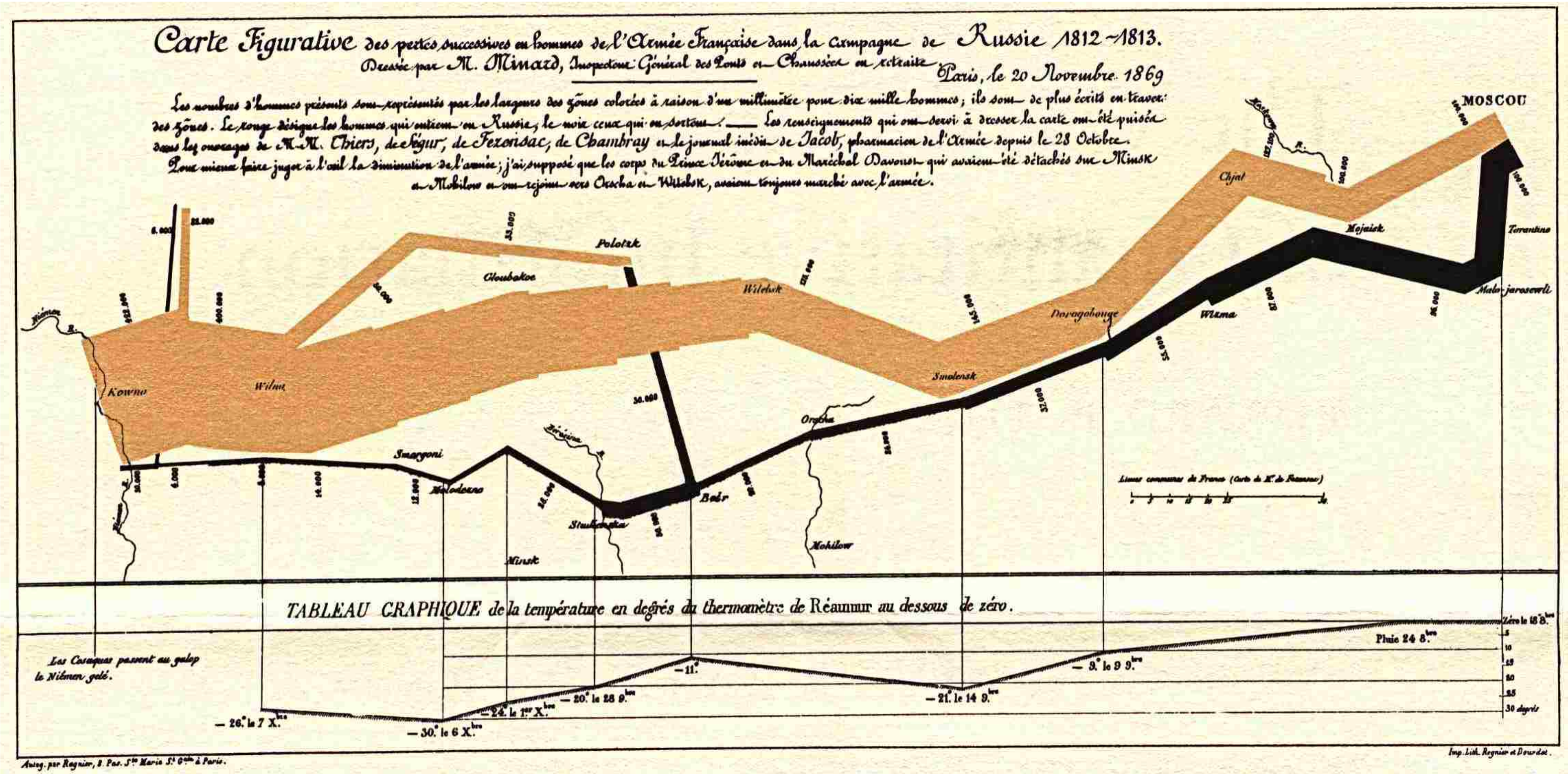
F. Giannotti& M. Nanni KDD Lab – ISTI – CNR
Pisa, Italy

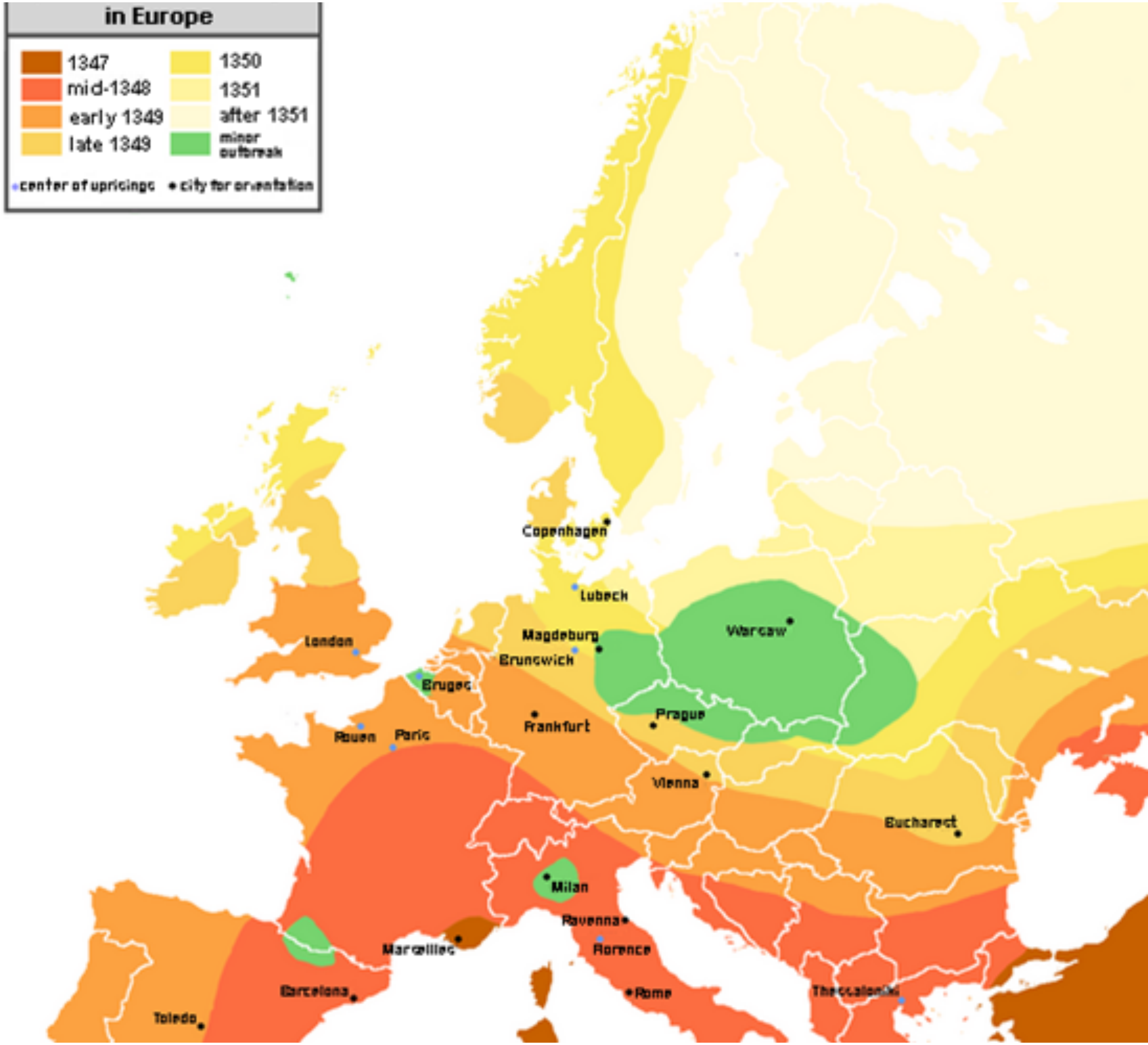
# Outline Mobility Data Mining

- Introduction
- MDM methods
  - Clustering
  - Trajectory Pattern Mining
  - Prediction
- MDM methods at work. Understanding Human Mobility
  - Dimensions of mobility analytics
  - Models of human mobility
  - The Mobility Atlas

- Module 3 Case studies
  - OD Matrix, D4D, Sociometer,
  - Network& Mobility

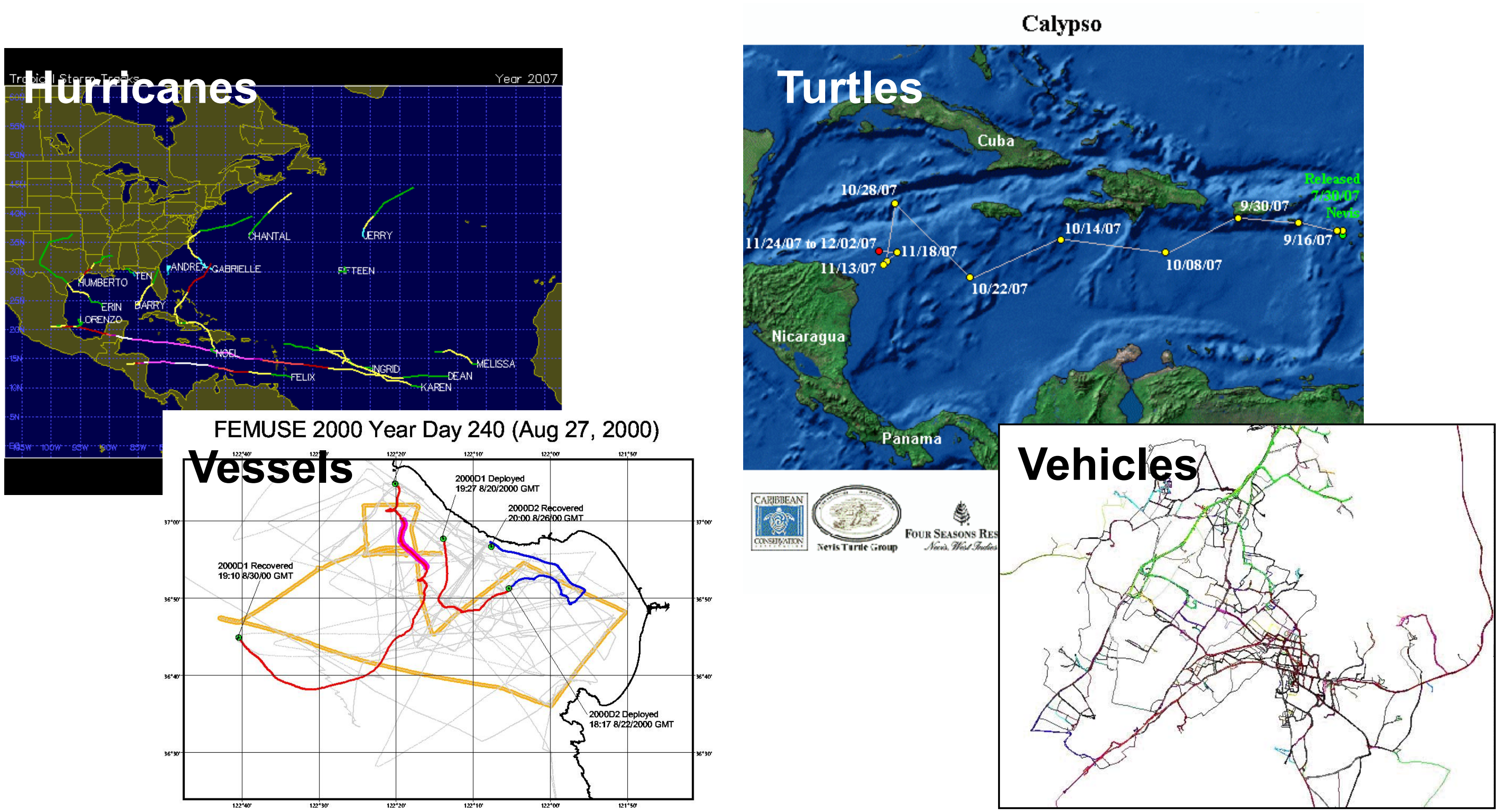
# Understanding Human Mobility: a long path




in Europe
1347
mid-1348
early 1349
late 1349
1350
1351
after 1351
minor outbreak
center of uprisings
city for orientation
Copenhagen
Lubeck
Magdeburg
Brunswick
Warsaw
London
Bruges
Rouen
Paris
Frankfurt
Prague
Vienna
Bucharest
Milan
Ravenna
Florence
Marseilles
Barcelona
Toledo
Rome
Thessaloniki

# Moving Object Data

- Several domains:



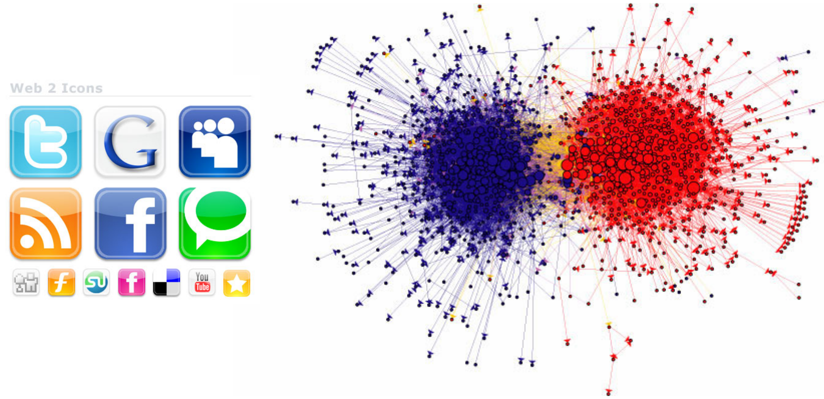
# Complexity of the Moving Object Data

- Uncertainty
  - Sampling rate could be inconstant: From every few seconds transmitting a signal to every few days transmitting one
  - Data can be sparse: A recorded location every 3 days
- Noise
  - Erroneous points (e.g., a point in the ocean)
- Background
  - Cars follow underlying road network
  - Animals movements relate to mountains, lakes, ...
- Movement interactions
  - Affected by nearby moving objects

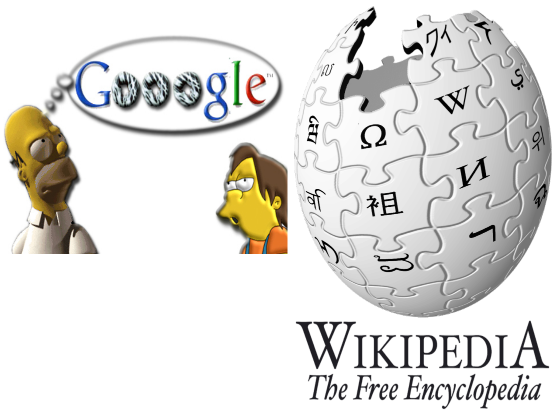
# The novelty : BIG DATA

What we buy

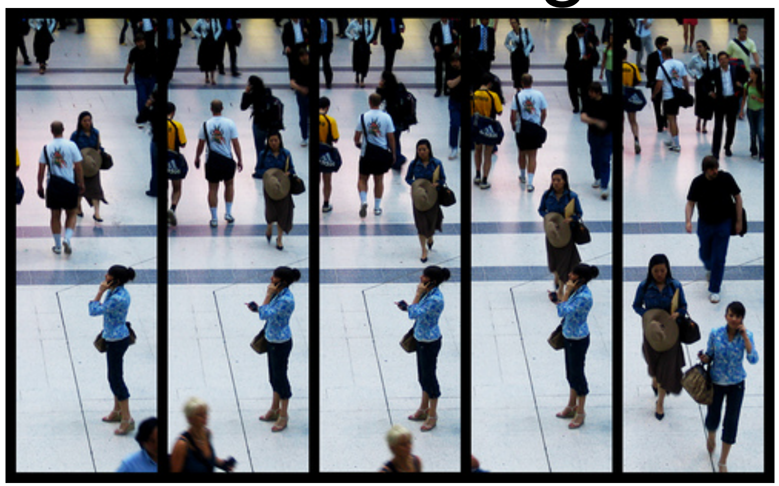
Whom we interact with




What we search for



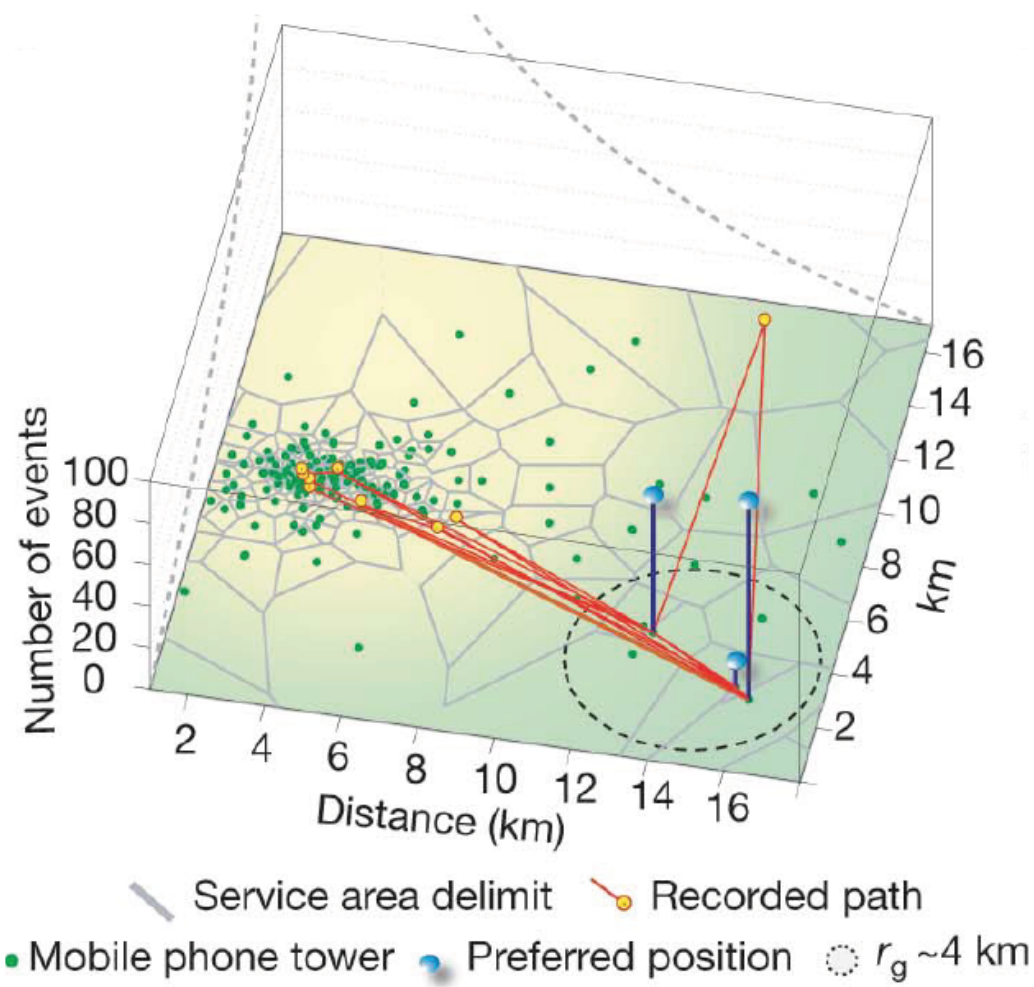

Where we go



# Why Mining Moving Object Data?

- Large diffusion of mobile devices, mobile services and location-based services

# Country-wide mobile phone data



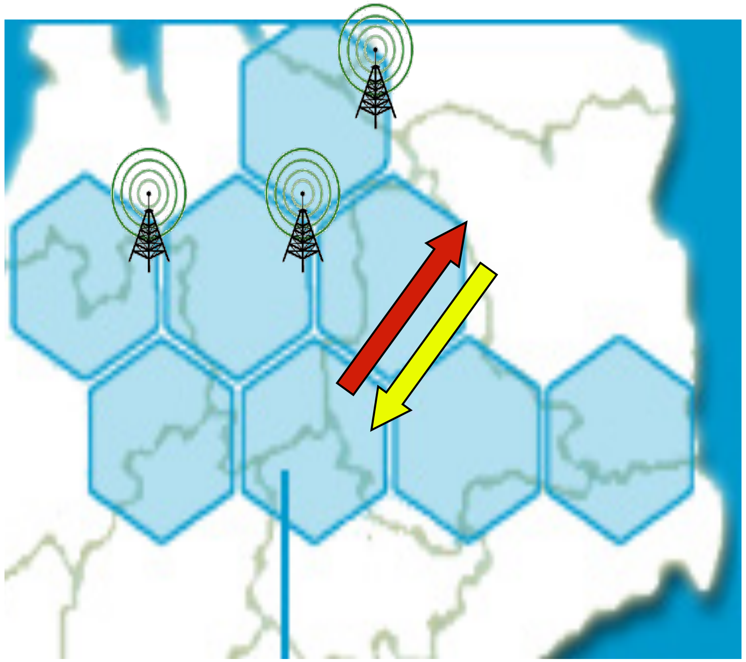


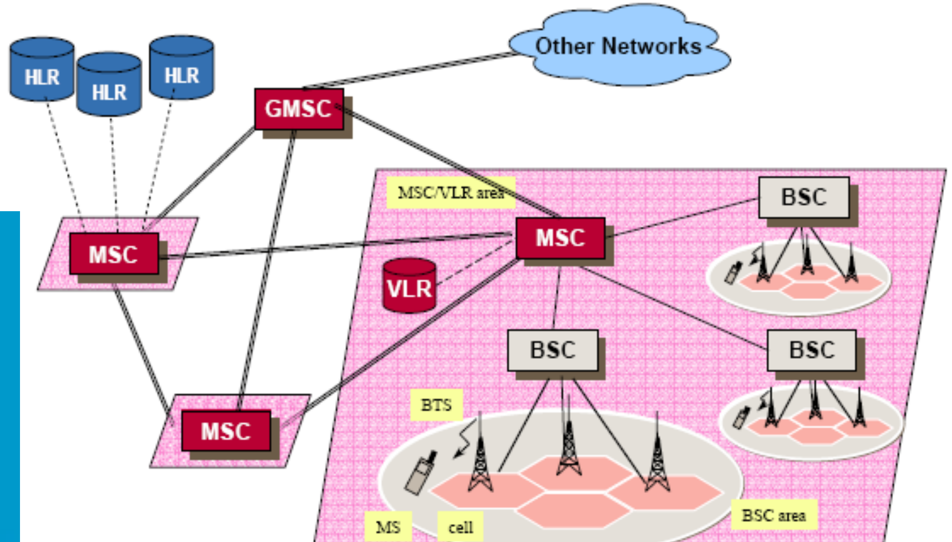
when you call

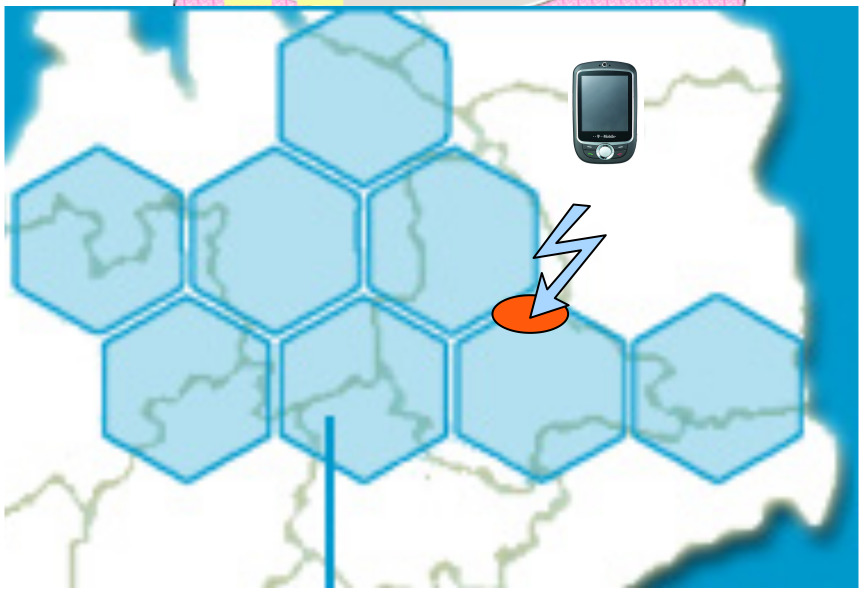
where you call

who you call

# GSM roaming CDR data –



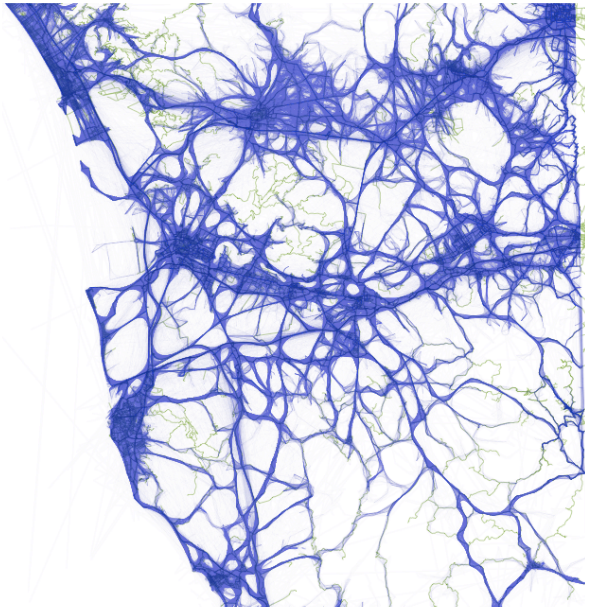


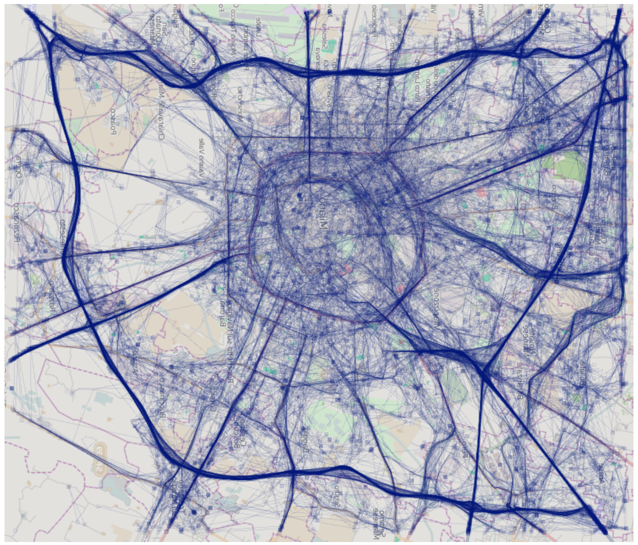



# GPS tracks

Onboard navigation devices send GPS tracks to central servers

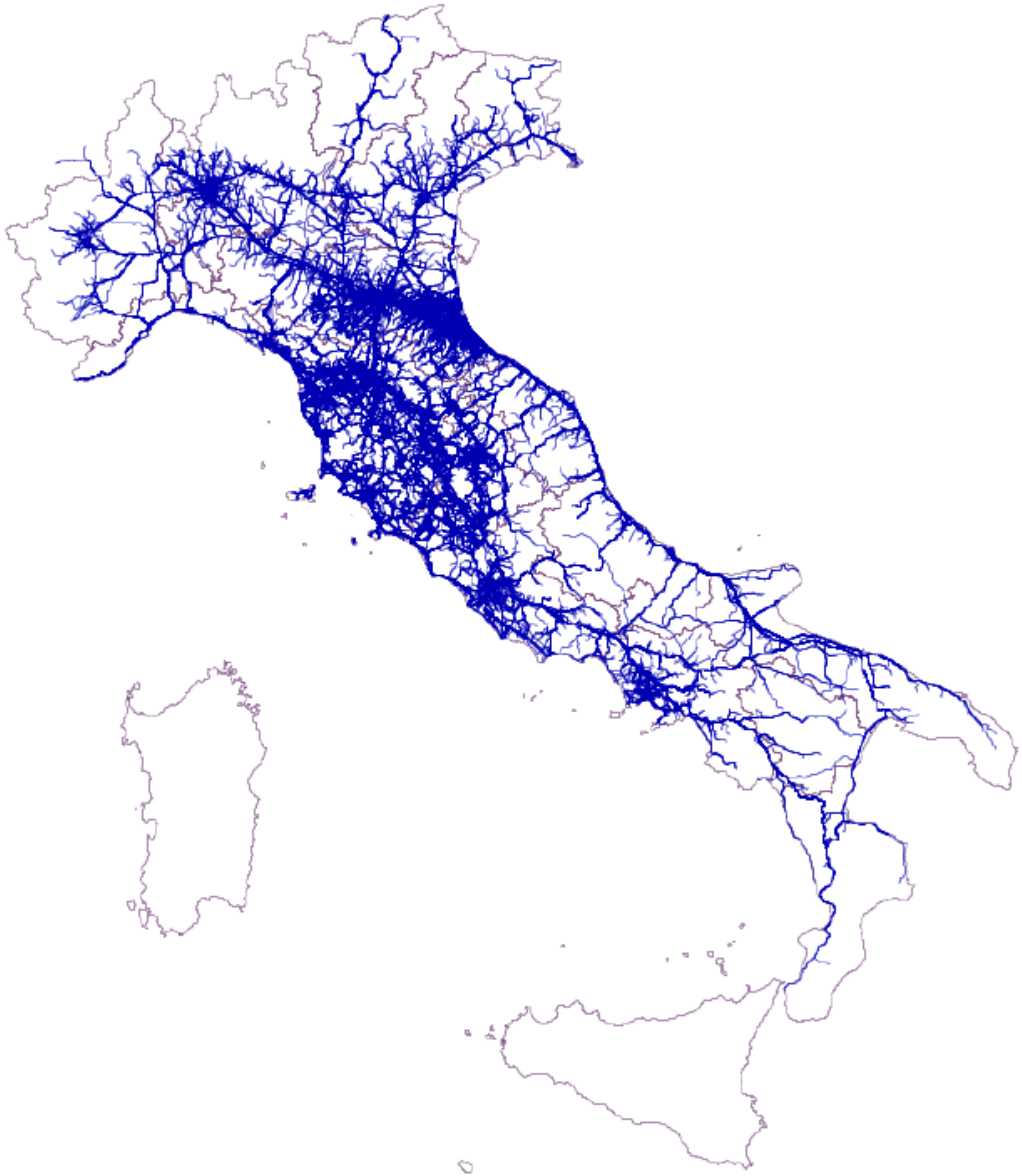
Ide;Time;Lat;Lon;Height;Course;Speed;PDOP;State;NSat
…
8;22/03/07 08:51:52;50.777132;7.205580; 67.6;345.4;21.817;3.8;1808;4
8;22/03/07 08:51:56;50.777352;7.205435; 68.4;35.6;14.223;3.8;1808;4
8;22/03/07 08:51:59;50.777415;7.205543; 68.3;112.7;25.298;3.8;1808;4
8;22/03/07 08:52:03;50.777317;7.205877; 68.8;119.8;32.447;3.8;1808;4
8;22/03/07 08:52:06;50.777185;7.206202; 68.1;124.1;30.058;3.8;1808;4
8;22/03/07 08:52:09;50.777057;7.206522; 67.9;117.7;34.003;3.8;1808;4
8;22/03/07 08:52:12;50.776925;7.206858; 66.9;117.5;37.151;3.8;1808;4
8;22/03/07 08:52:15;50.776813;7.207263; 67.0;99.2;39.188;3.8;1808;4
8;22/03/07 08:52:18;50.776780;7.207745; 68.8;90.6;41.170;3.8;1808;4
8;22/03/07 08:52:21;50.776803;7.208262; 71.1;82.0;35.058;3.8;1808;4
8;22/03/07 08:52:24;50.776832;7.208682; 68.6;117.1;11.371;3.8;1808;4
…



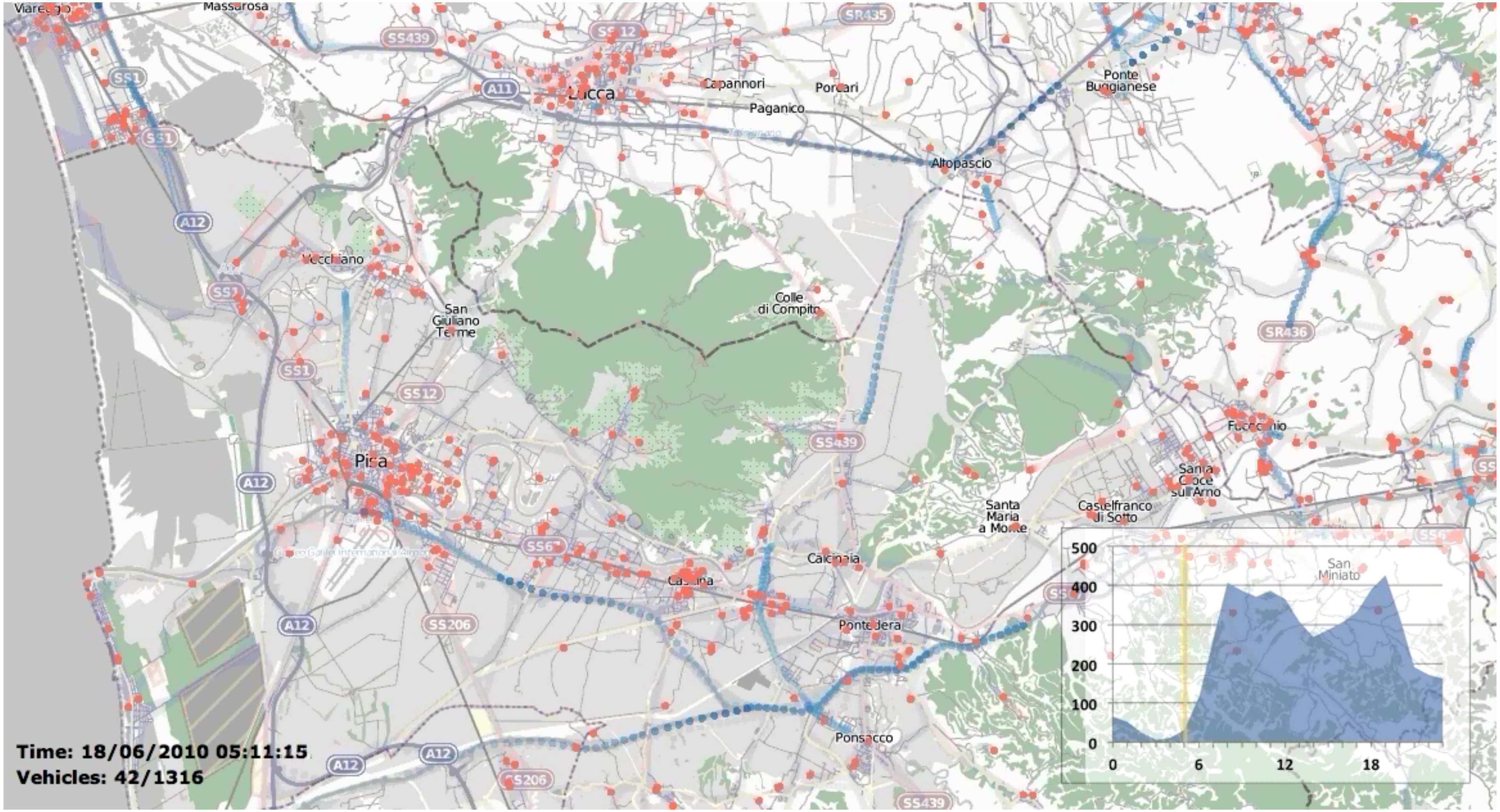


Sampling rate ~3 secs

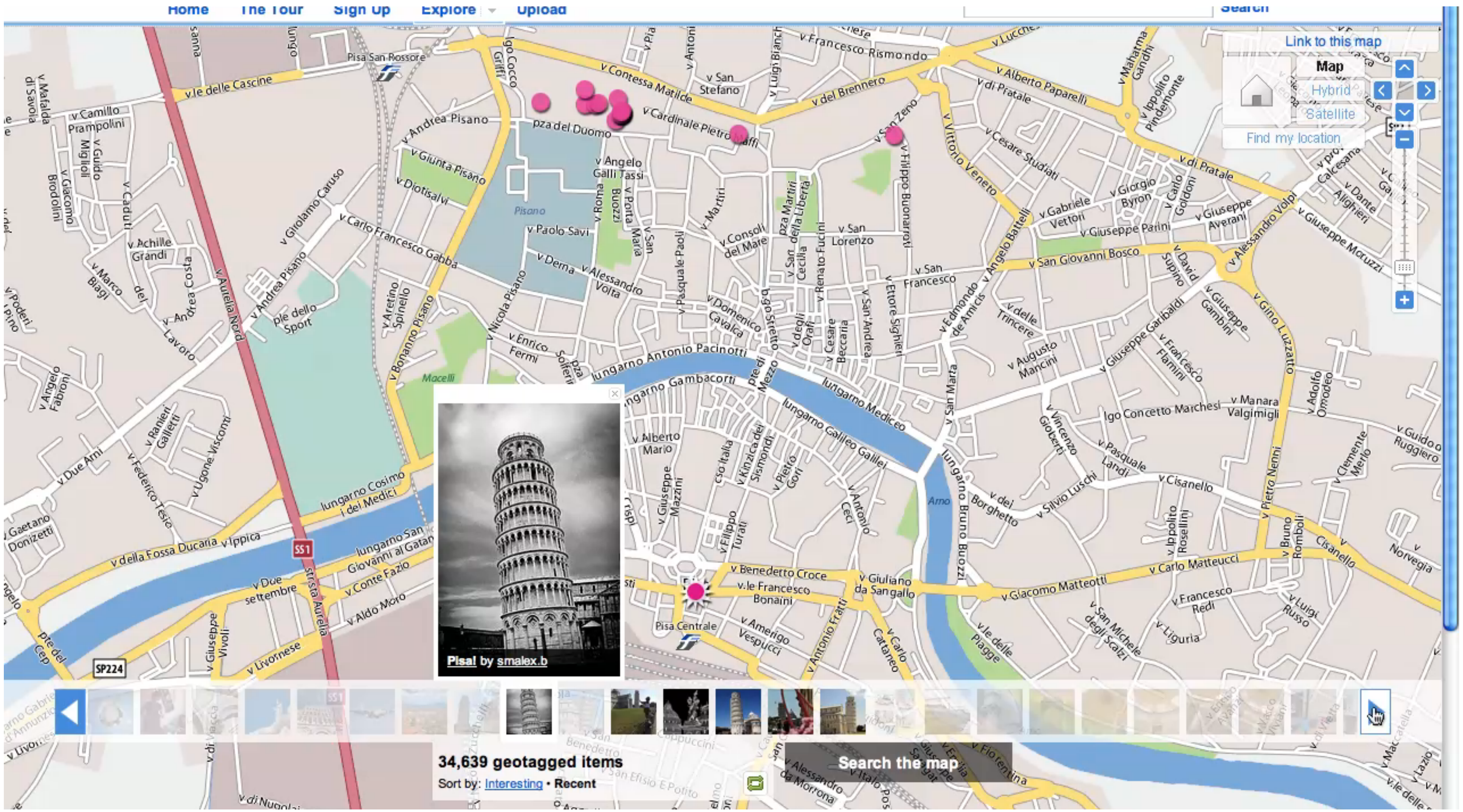
Spatial precision ~ 10 m



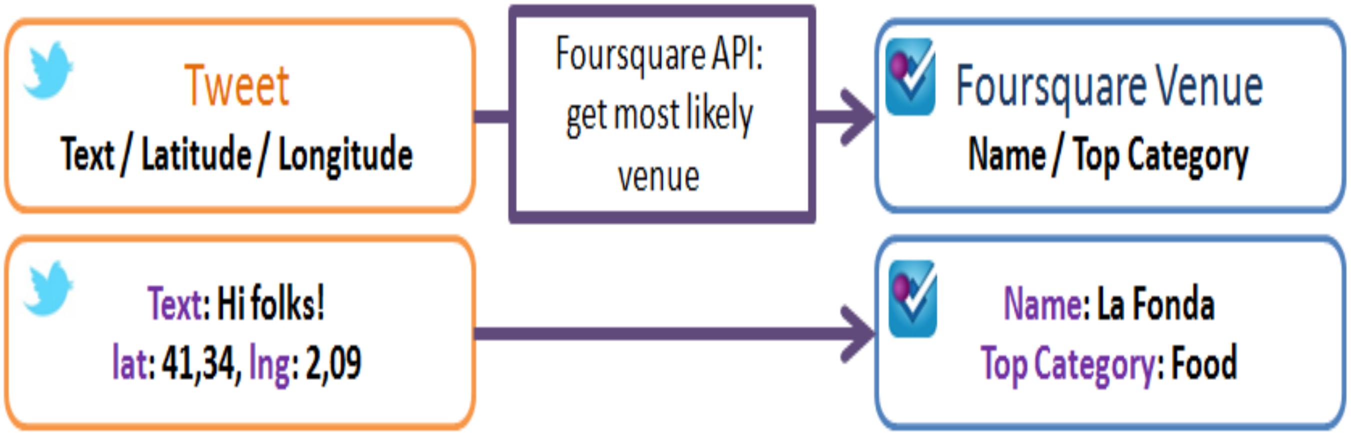
# Urban Mobility Complexity: vehicles



# Social networks



# Tweeter



Tweet
Text / Latitude / Longitude
Foursquare API: get most likely venue
Foursquare Venue
Name / Top Category
Text: Hi folks!
lat: 41,34, lng: 2,09
Name: La Fonda
Top Category: Food

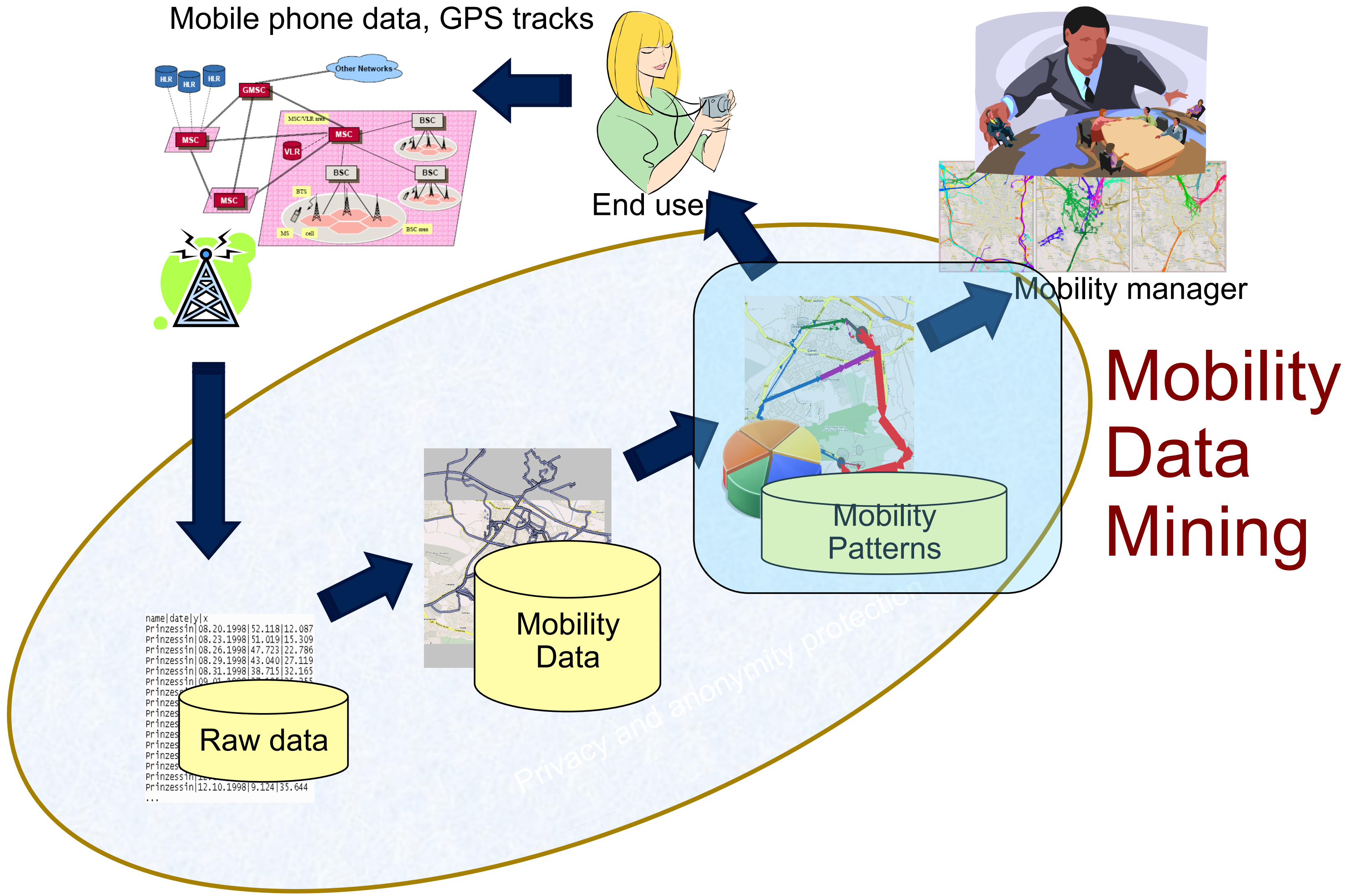
# Research Impacts

- Moving object and trajectory data mining has many important, real-world applications driven by the real need
  - Ecological analysis (e.g., animal scientists)
  - Weather forecast
  - Traffic control
  - Location-based services
  - Homeland security (*e.g.*, border monitoring)
  - Law enforcement (*e.g.*, video surveillance)
  - ...

# outline

- Introduction
- MDM methods
  - Clustering
  - Trajectory Pattern Mining
  - Prediction
  - Semantic enrichment
- MDM methods at work. Understanding Human Mobility
  - Dimensions of mobility analytics
  - Models of human mobility
  - The Mobility Atlas
- Module 3 Case studies
  - OD Matrix, D4D, Sociometer, Correlation Patterns,
  - Network& Mobility

# The (GeoP)KDD process

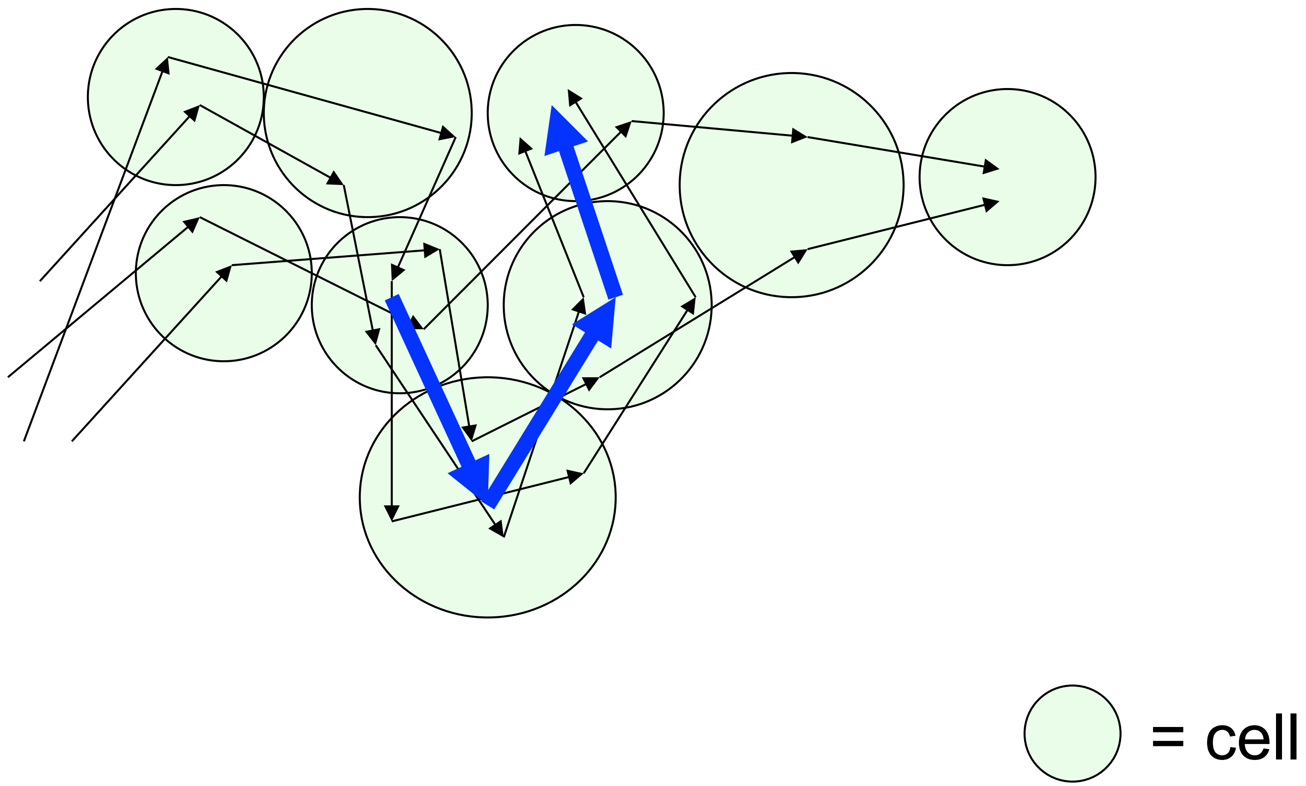


# Data mining …

- … is about finding models that emerge directly from the data
  - Data-driven vs hypothesis-driven analysis
- Local models
  - **Patterns**: find groups of items/events that frequently co-occur in the data
- Global models
  - **Clustering**: find a natural partition of the data into groups of similar objects
  - **Classification**: find a function that predicts the value of a specified variable given the values of the others

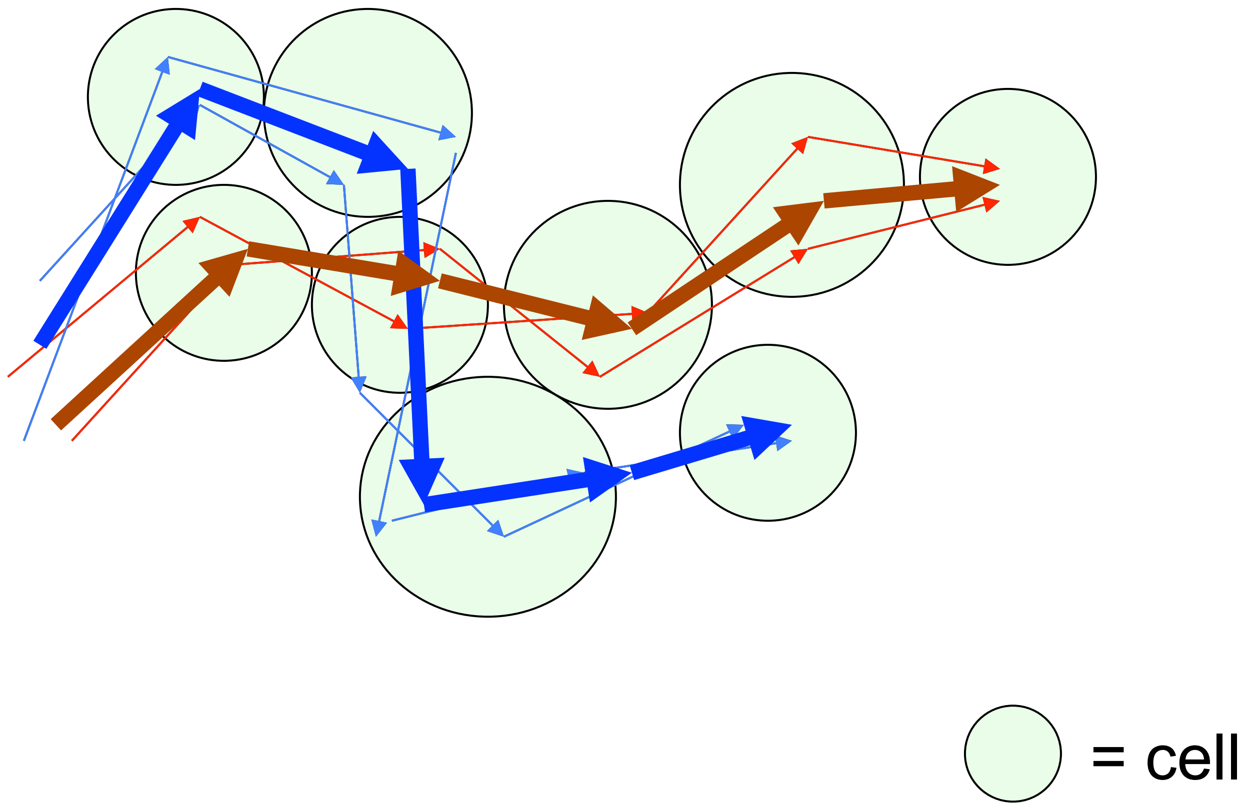
# Trajectory patterns

- Discover frequently followed itineraries



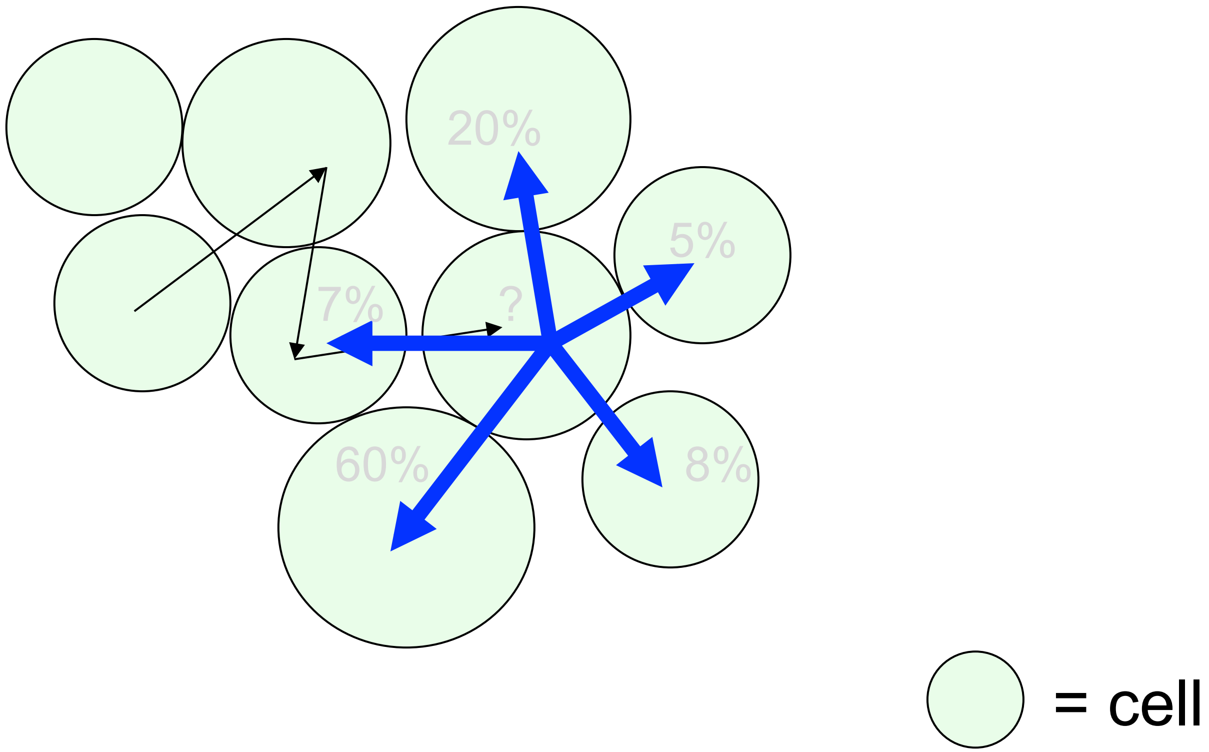
# Trajectory Clustering

- Group together similar trajectories
- For each group produce a summary



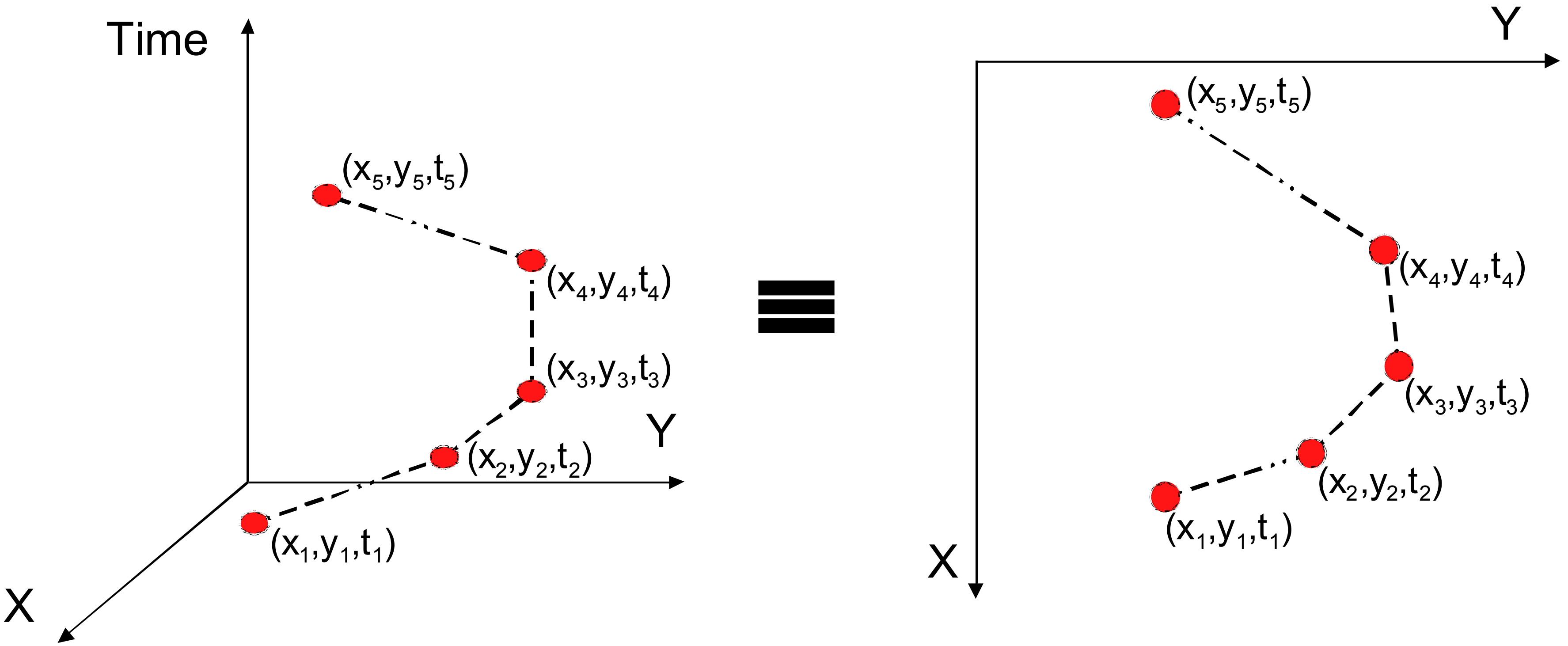
# Trajectory classification and prediction

- Extract behaviour rules from history
- Use rules to predict behaviour of future users

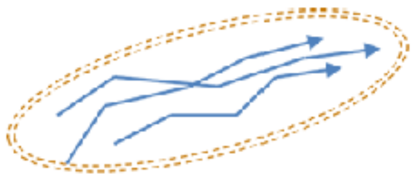


# Trajectory data

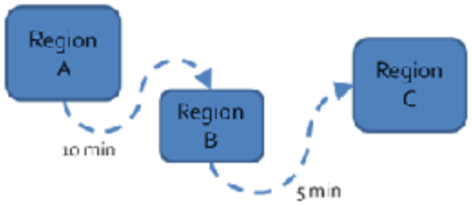
- Mobility of an object is described by a set of trips
- Each trip is a trajectory, i.e. a sequence of time-stamped locations



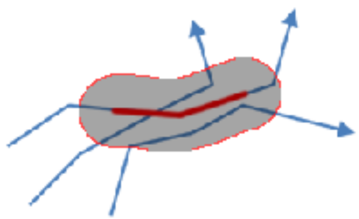
# Basic mobility patterns and models



- T-Cluster: represents a group of similar trajectories




- T-Pattern: represents trajectory segments that visit the same sequence of regions with similar transition times



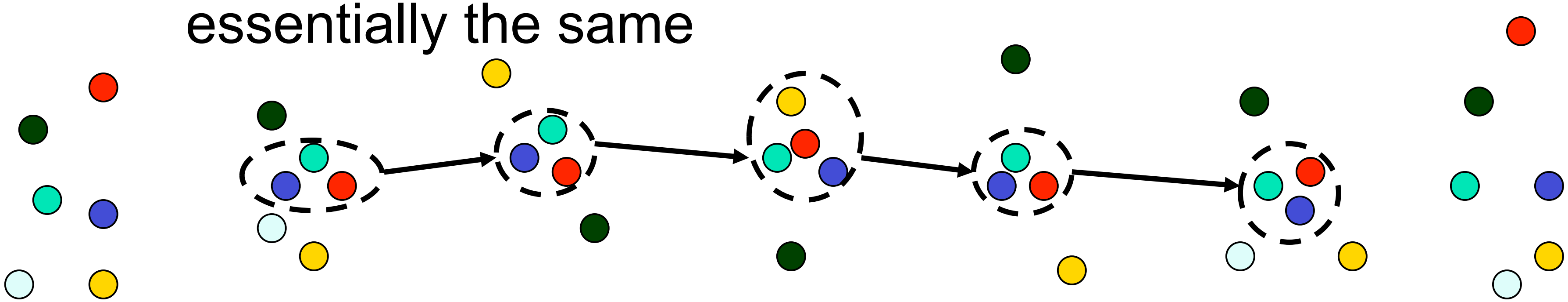
- T-Flock: represents trajectory segments that move together for a time interval

# Trajectory patterns

Are there groups of objects that move together for some time?

# Moving Clusters

- A ***moving cluster*** is a set of objects that move close to each other for a long time interval
  - **Note**: Moving clusters and flock patterns (see later) are essentially the same



- Formal Definition [Kalnis et al., SSTD'05]:
  - A ***moving cluster*** is a sequence of (snapshot) clusters $\boldsymbol{c_1, c_2, \ldots, c_k}$ such that for each timestamp $i$ $(1 \leq i < k)$, $|c_i \cap c_{i+1}| / |c_i \cup c_{i+1}| \geq \theta \quad (0 < \theta \leq 1)$

# Relative Motion Patterns

**(Laube et al. 04, Gudmundsson et al. 07)**

- ***Flock:*** At least *m* entities are within a circular region of **radius *r*** and they move in the same direction

- ***Leadership*:** At least *m* entities are within a circular region of radius *r*, they move in the same direction, and **at least one of the entities was already heading in this direction for at least *s* time steps**

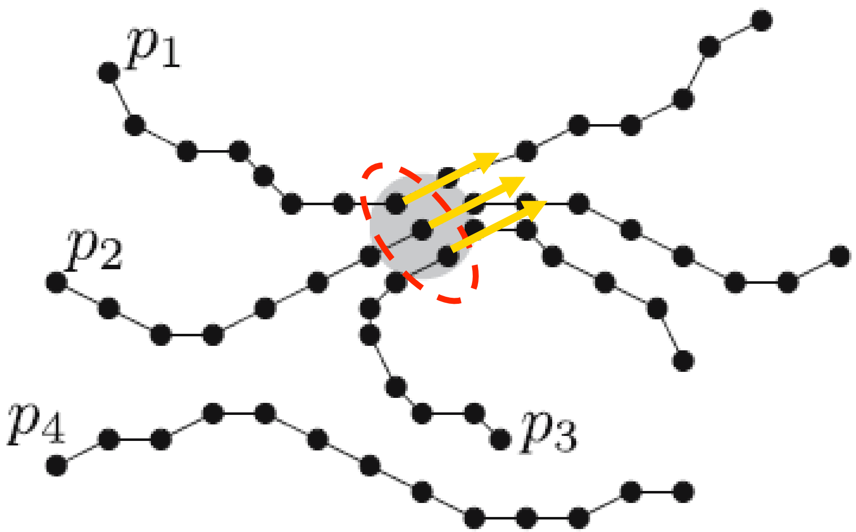
- ***Convergence:*** At least *m* entities will **pass through** the same circular region of radius *r* (assuming they keep their direction)

- ***Encounter:*** At least *m* entities will be **simultaneously inside** the same circular region of radius *r* (assuming they keep their speed and direction)

# Relative Motion Patterns

**(Laube et al. 04, Gudmundsson et al. 07)**

- ***Flock*** ($m$ > 1, $r$ > 0): At least $m$ entities are within a circular region of **radius *r*** and they move in the same direction

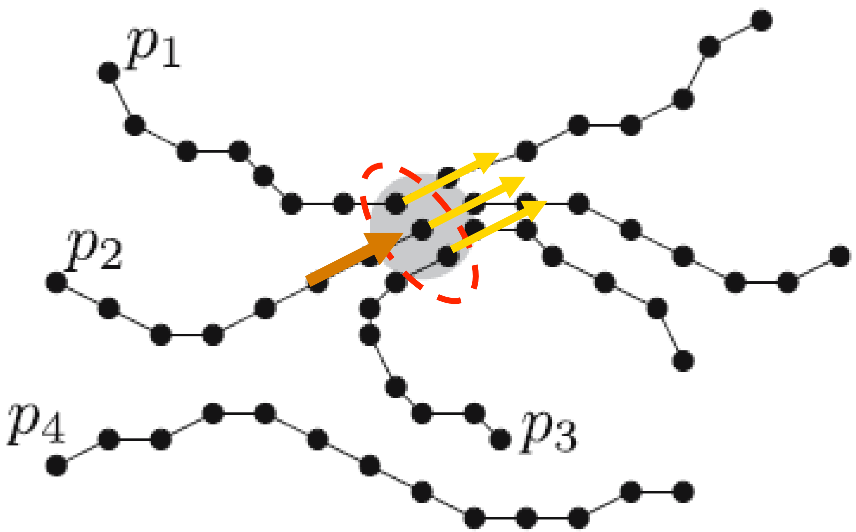



An example of a ***flock*** pattern for $p_1$, $p_2$, and $p_3$ at 8th time step; also a ***leadership*** pattern with $p_2$ as the leader

# Relative Motion Patterns

**(Laube et al. 04, Gudmundsson et al. 07)**

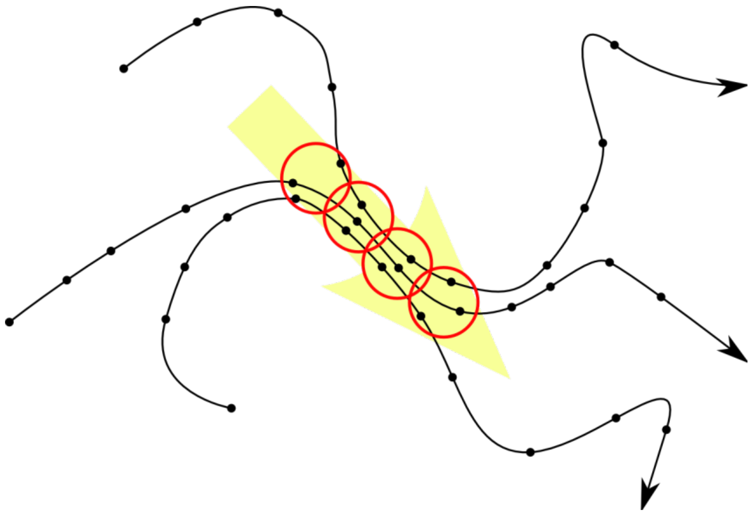
- ***Leadership*** (*m* > 1, *r* > 0, *s* > 0) At least *m* entities are within a circular region of radius *r*, they move in the same direction, and **at least one of the entities was already heading in this direction for at least *s* time steps**




An example of ***leadership*** pattern with $p_2$ as the leader

# Basic mobility patterns & models: T-Flocks

- Group of objects that move together (close to each other) for a time interval

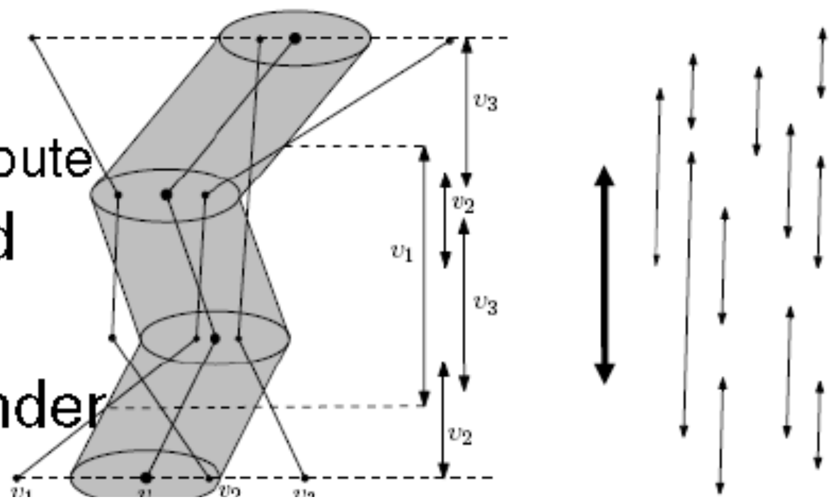


M. Wachowicz, R. Ong, C. Renso, M. Nanni: **Finding moving flock patterns among pedestrians through collective coherence.** International Journal of Geographical Information Science 25(11): 1849-1864 (2011)

# Computing Flock Patterns

- *Approximate* flocks
  - Convert overlapping segments of length *k* to points in a *2k*-dimensional space
  - Find *2k*-d pipes that contain at least *m* points

- *Longest duration* flocks
  - For every entity *v*, compute a cylindrical region and the intervals from the intersection of the cylinder
  - Pick the longest one



# Convoy: An Extension of Flock Pattern

**(Jeung et al. ICDE'08 & VLDB'08)**

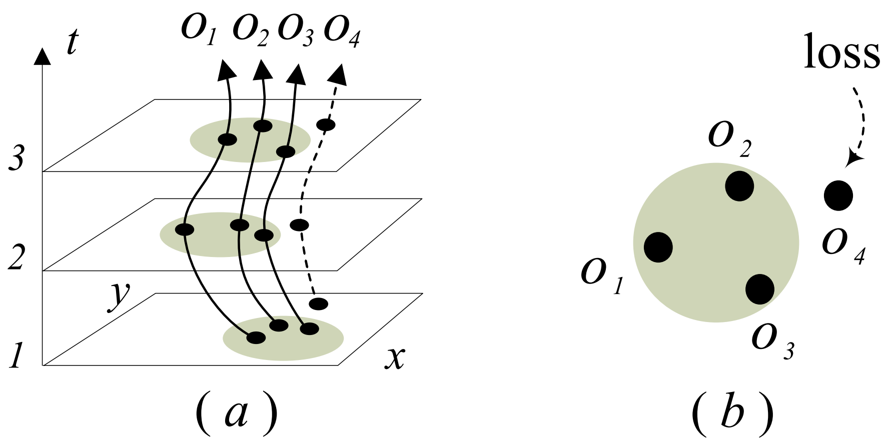



**Figure 1:** ***Lossy-flock*** **Problem**

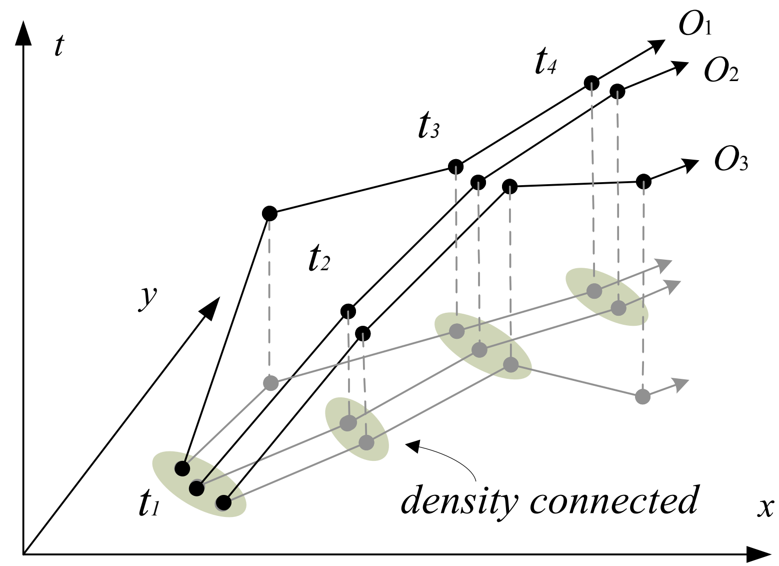



**Figure 4: An Example of a Convoy**

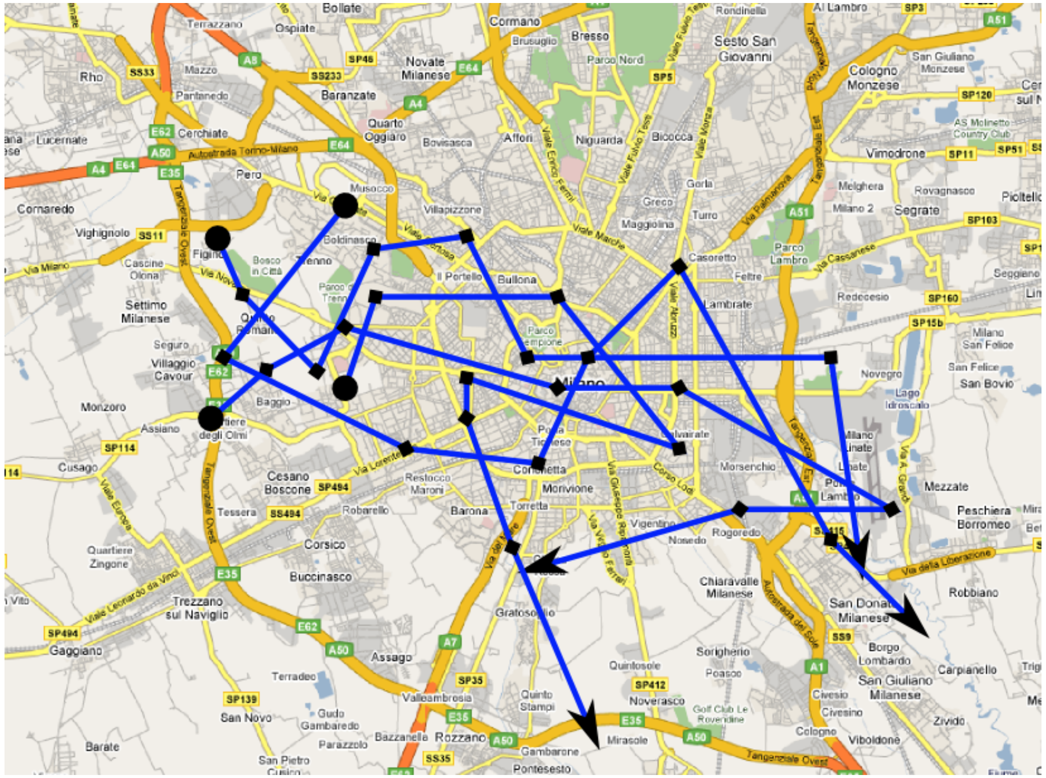
- Flock pattern has rigid definition with a circle
- Convoy use *density-based clustering* at each timestamp

# Trajectory pattern mining

- Are there groups of objects that perform a sequence of movements, with
- similar timings though possibly during completely dierent moments?

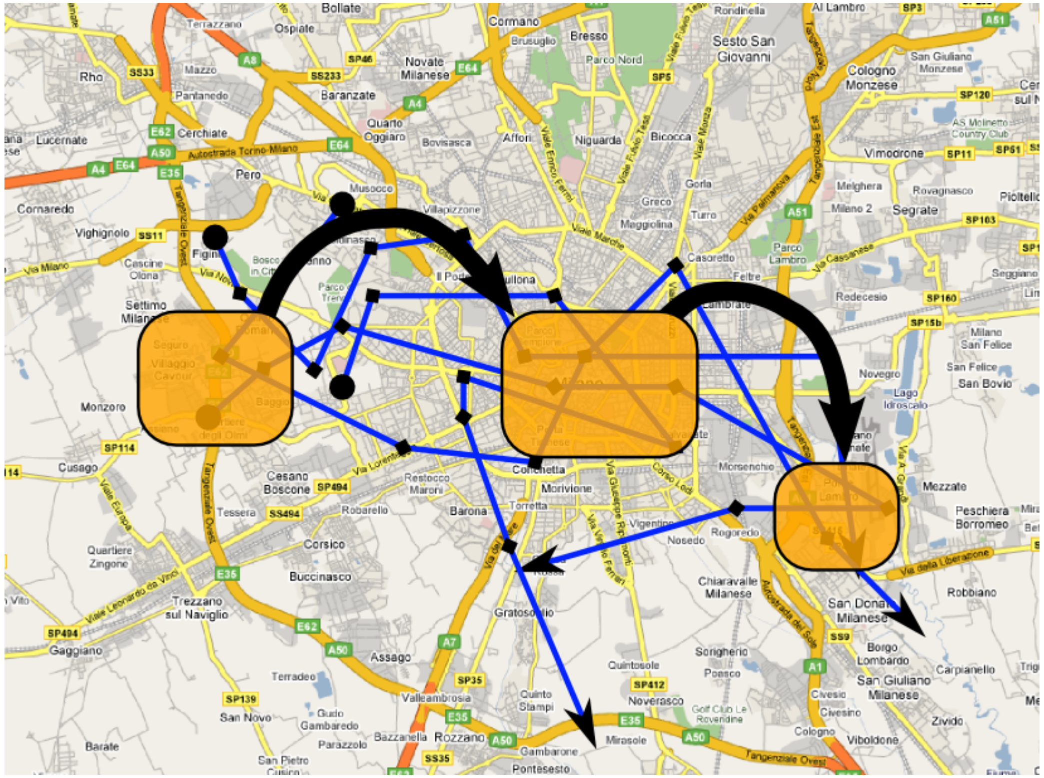
# Q: What is a trajectory pattern?





# A: A spatio-temporal sequential pattern

- A sequence of visited regions, **frequently** visited in the **specified order** with **similar transition times**

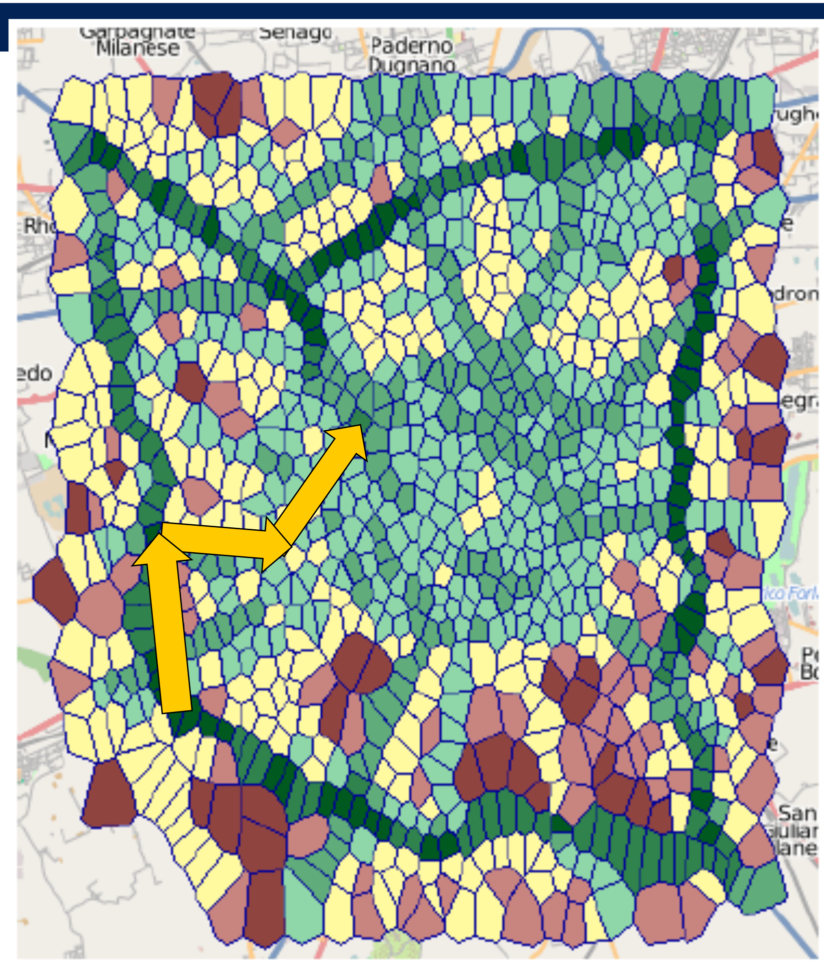



■

# Simpler case: GSM trajectories

- Each trajectory in the database is a time-stamped sequence of **predefined areas** (antennas)

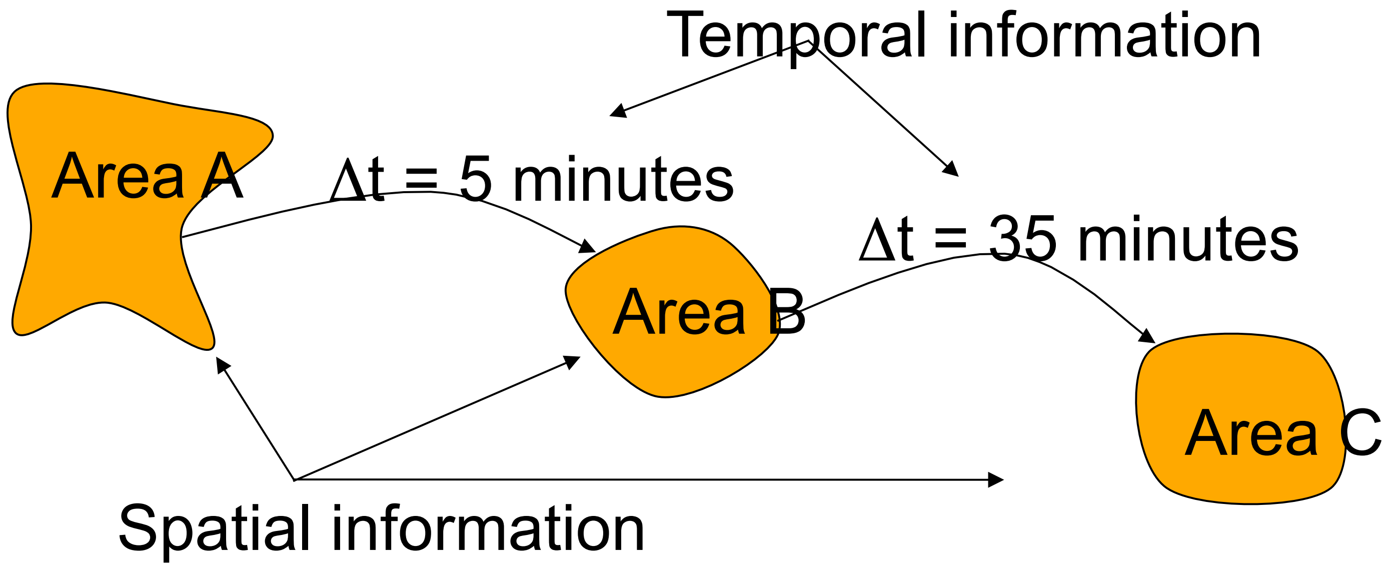
- A T-pattern is a sequence of such areas that occurs often in the data with similar travel time



# Trajectory pattern

$$A_0 \xrightarrow{t_1} A_1 \xrightarrow{t_2} \dots \quad A_{n-1} \xrightarrow{t_n} A_n$$

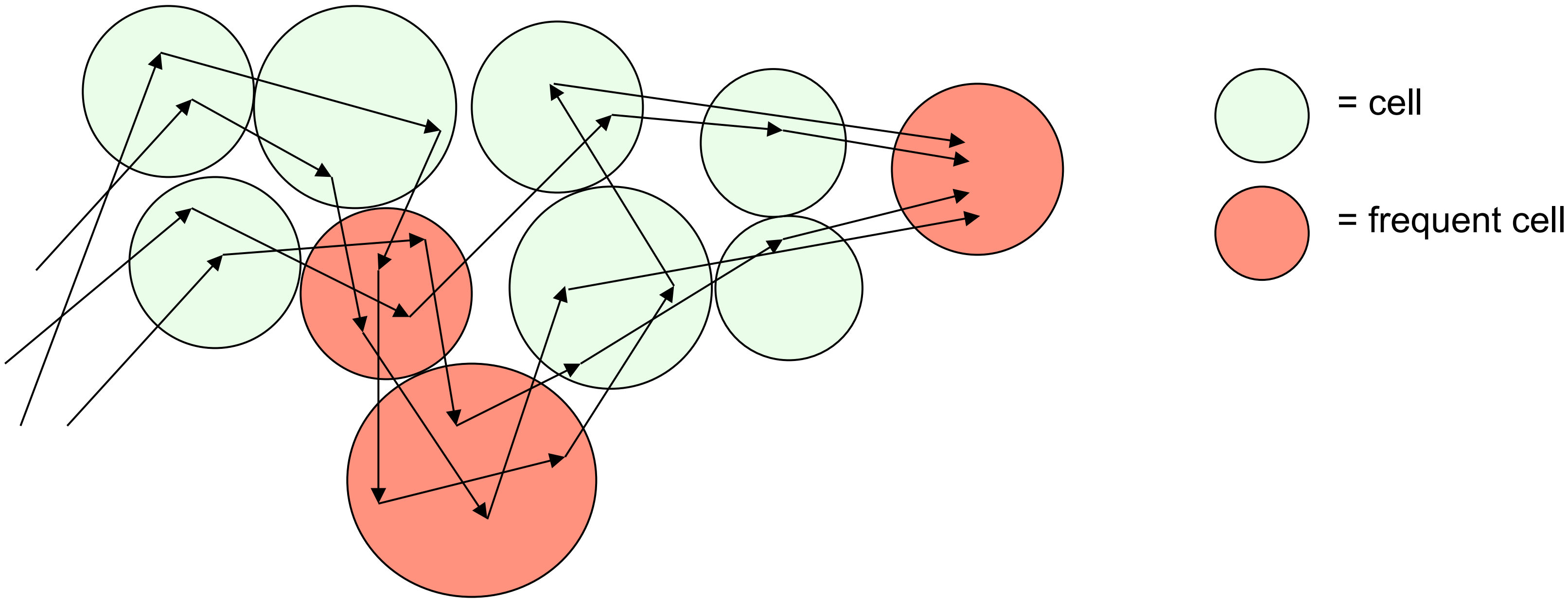
- $t_i$ = transition time




$$\mathit{Station} \xrightarrow{20\text{min.}} \mathit{Castle} \xrightarrow{65\text{min.}} \mathit{Museum}$$

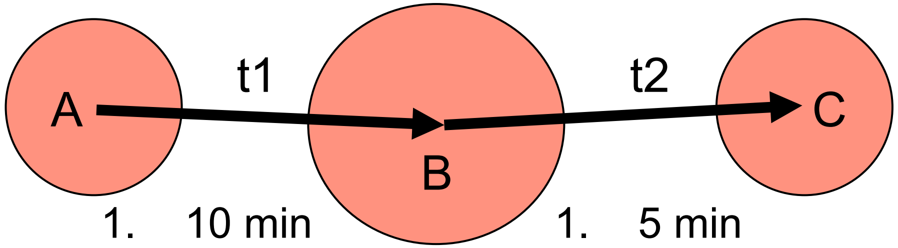
# Spatial component

Discover frequent cells using a minimum support threshold (4)



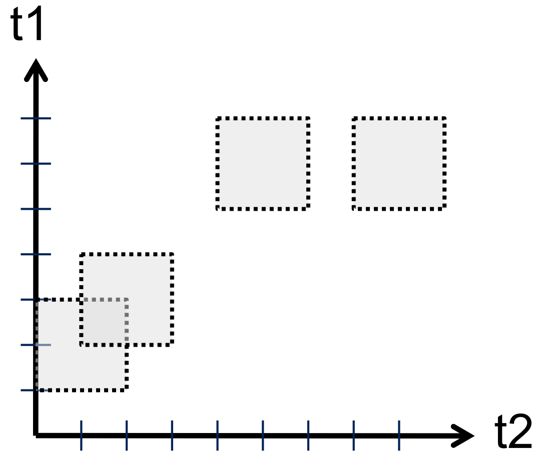
# Temporal component

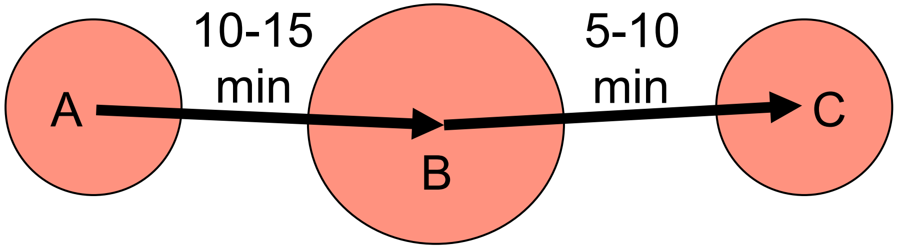
Discover frequent time interval using a tollerance (5 min):




1. 10 min
2. 15 min
3. 30 min
4. 30 min

1. 5 min
2. 10 min
3. 25 min
4. 40 min

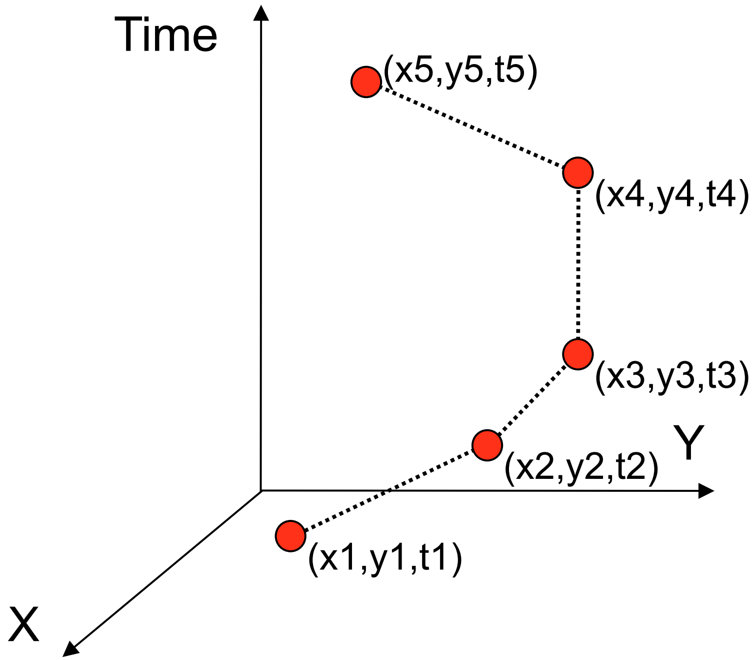


# More complex case: GPS trajectories

- Each trajectory in the database is a time-stamped sequence of points: $<(x_1,y_1,t_1), \ldots, (x_n,y_n,t_n)>$

- In general, no predefined regions/ areas



# T-Patterns for trajectories



- A **Trajectory Pattern** (T-pattern) is a pair (**s**, $\alpha$):
  - **s** = $<(x_0,y_0),\ldots, (x_k,y_k)>$ is a sequence of k+1 locations
  - $\alpha$ = $<\alpha_1,\ldots, \alpha_k>$ are the transition times (*annotations*)

  also written as:

$$(x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1) \xrightarrow{\alpha_2} \cdots \xrightarrow{\alpha_k} (x_k, y_k)$$

- A trajectory t **supports** a T-pattern $T_p$ if t contains a sub-sequence S such that:
  - each $(x_i,y_i)$ in $T_p$ matches a point $(x_i',y_i')$ in S, and
  - the transition times in $T_p$ are similar to those in S

# Continuity issues (space & time)



- What does “matches” mean in space/time?
  - The same exact spatial location (x,y) usually never occurs twice
  - The same exact transition times usually do not occur twice

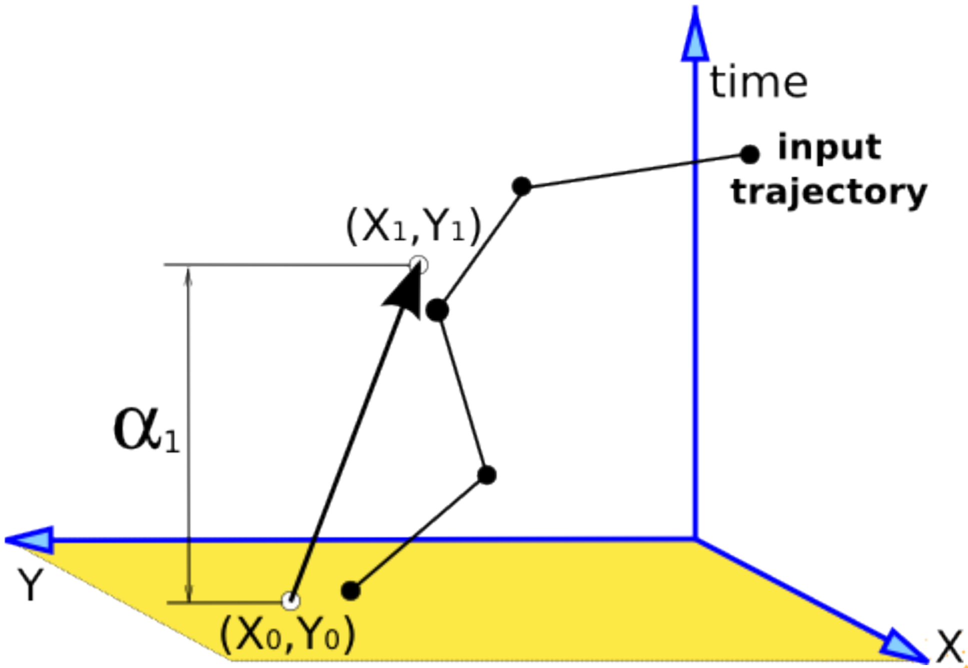
- Solution: allow approximation
  - a notion of *spatial neighborhood*
  - a notion of *temporal tolerance*

# T-Pattern: approximate occurrence



- Two points match if one falls within a **spatial neighborhood N()** of the other
- Two transition times match if their **temporal difference is ≤ τ**
- Example:

$$(x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1)$$

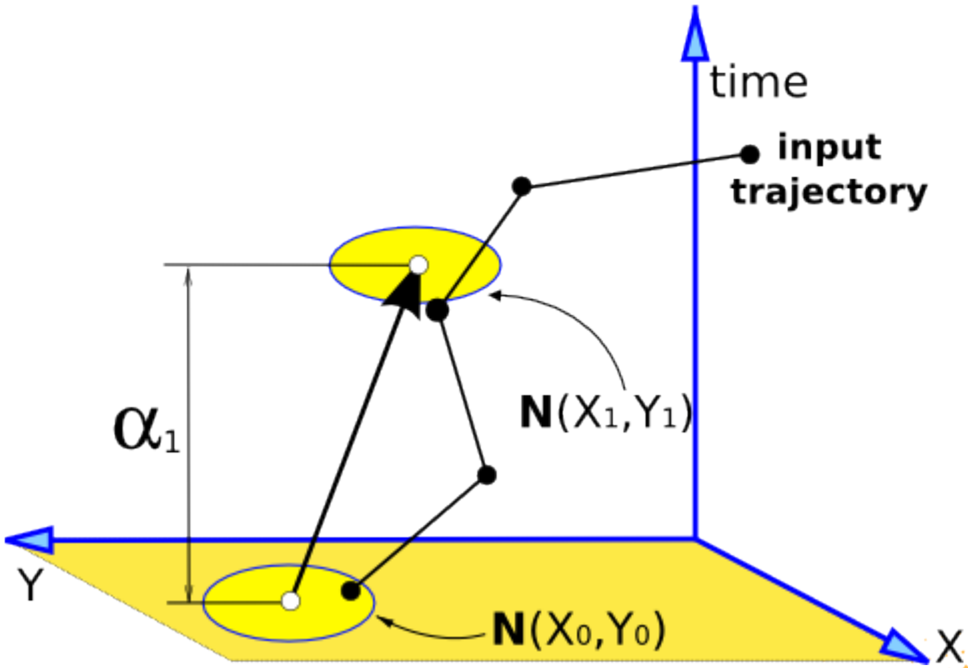


# T-Pattern: approximate occurrence



- Two points match if one falls within a **spatial neighborhood N()** of the other
- Two transition times match if their **temporal difference is ≤ τ**
- Example:

$$(x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1)$$

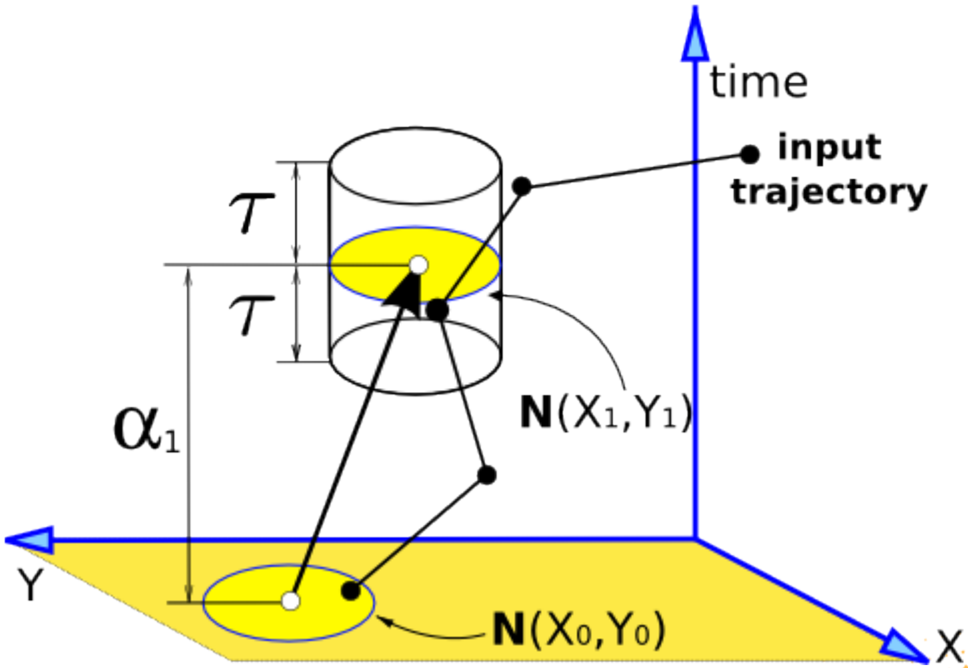


# T-Pattern: approximate occurrence

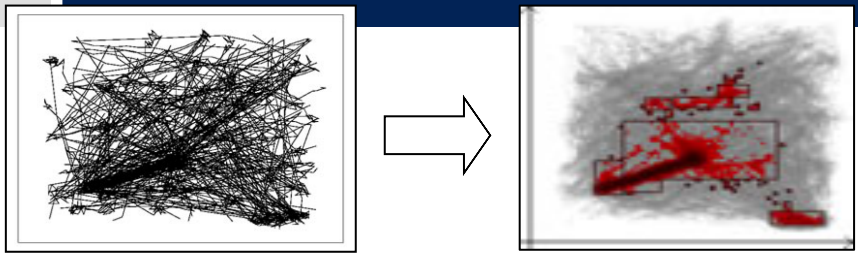


- Two points match if one falls within a **spatial neighborhood N()** of the other
- Two transition times match if their **temporal difference is ≤ τ**
- Example:

$$(x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1)$$

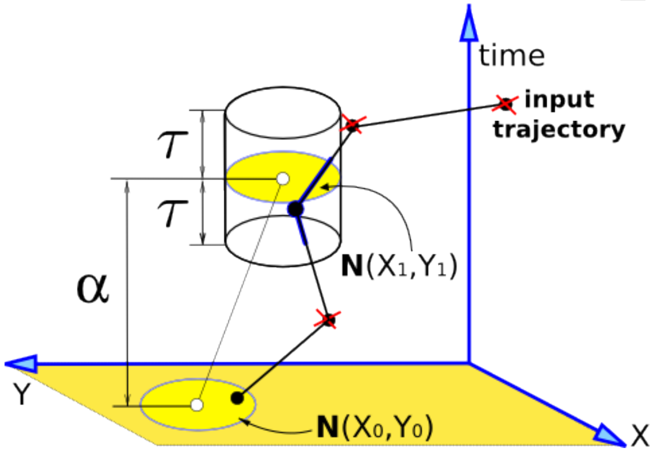


# T-Pattern discovery

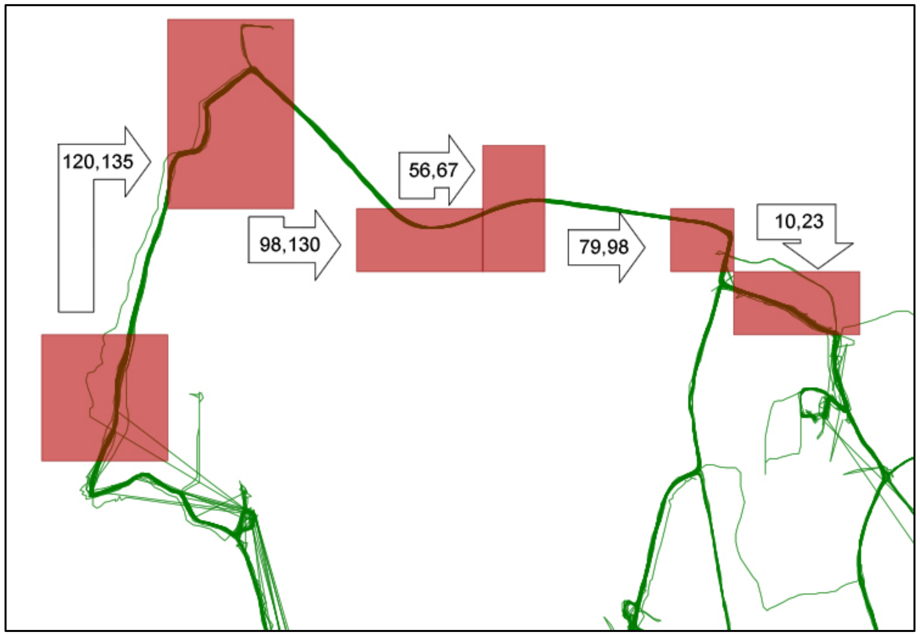


1- Find Regions of Interest

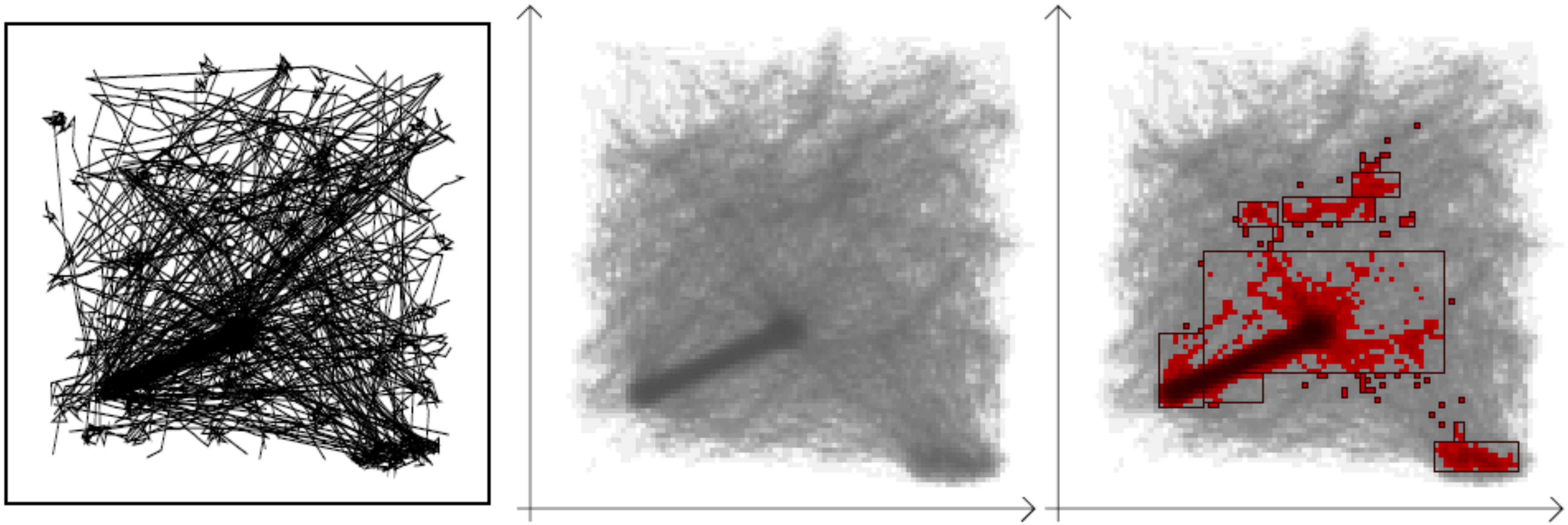
2- Find similar Trajectory in space and time




3- Extract patterns:



# Finding regions
## A usage-based heuristic

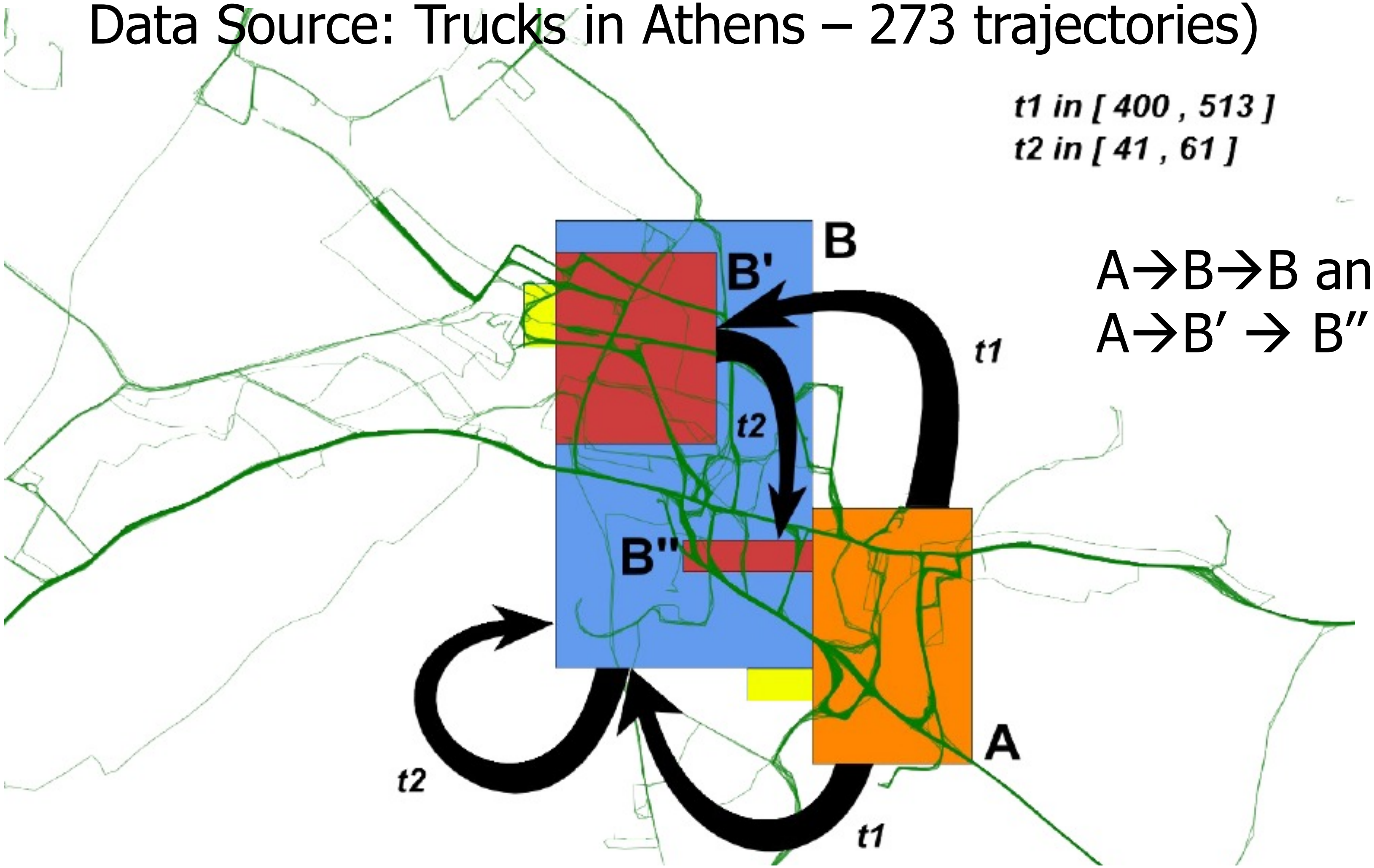


(a) input trajectories (b) density distribution (c) dense cells and extracted RoI

1. Impose a regular grid over space
2. Find dense cells (i.e., touched by many trajs.)
3. Coalesce cells into rectangles of bounded size

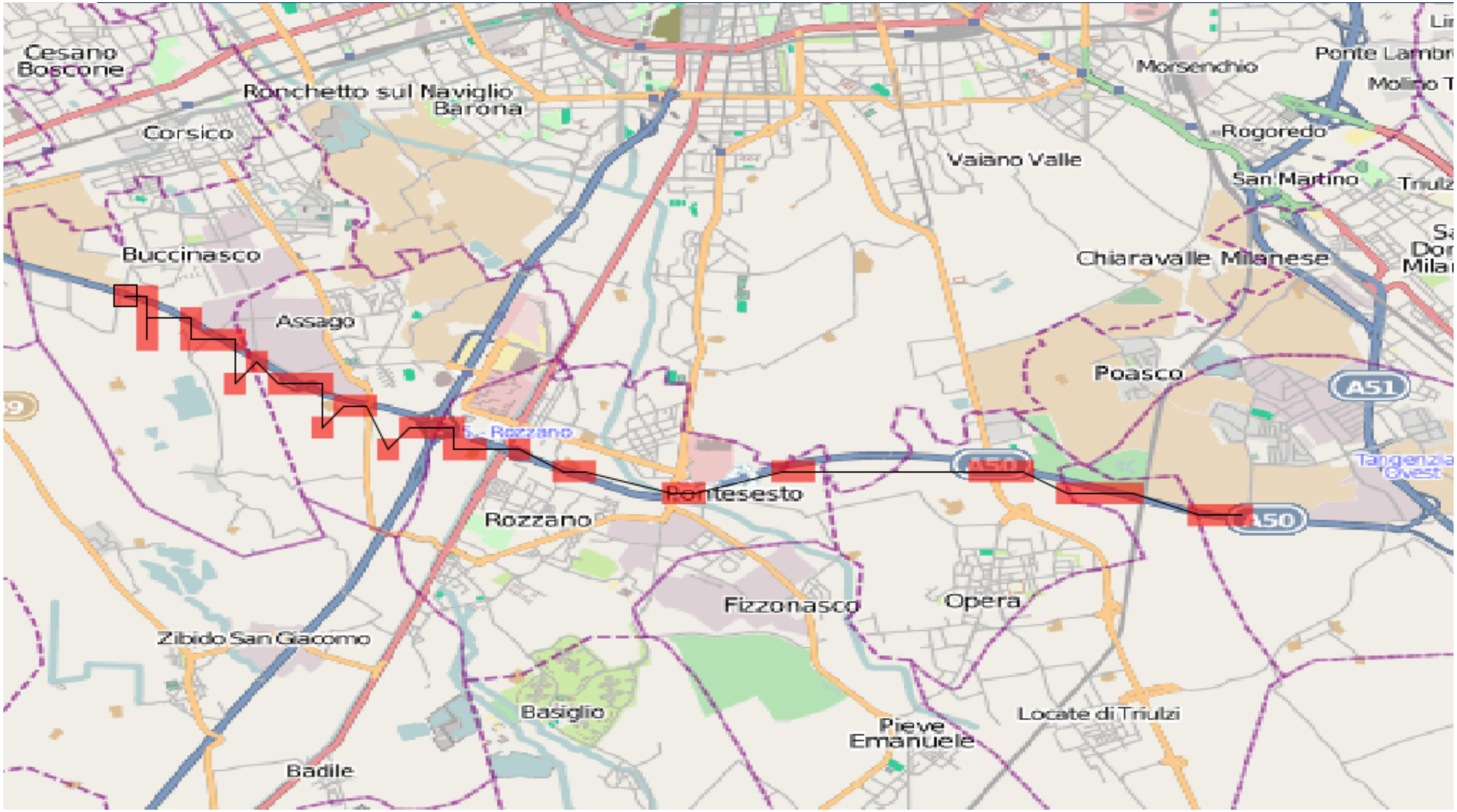
# Sample Trajectory-Patterns

Data Source: Trucks in Athens – 273 trajectories)




A→B→B and
A→B′ → B″

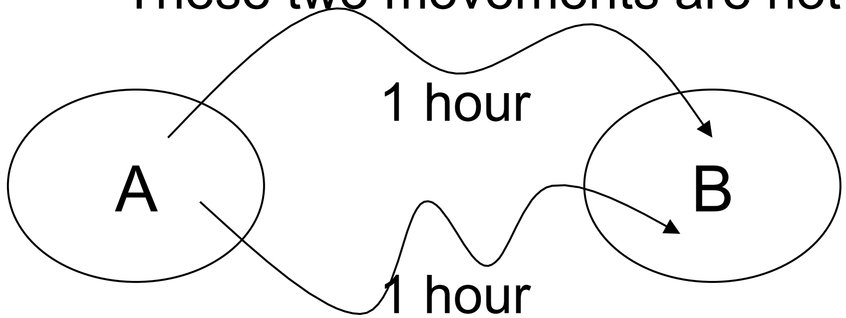
# A T-pattern


Cesano Boscone
Ronchetto sul Naviglio
Barona
Corsico
Morsenchio
Ponte Lambro
Rogoredo
Vaiano Valle
San Martino
Buccinasco
Chiaravalle Milanese
Assago
Poasco
A51
Rozzano
Pontesesto
A50
Tangenziale Ovest
Fizzonasco
Opera
Zibido San Giacomo
Basiglio
Locate di Triulzi
Pieve Emanuele
Badile

# Characteristics of Trajectory-Patterns

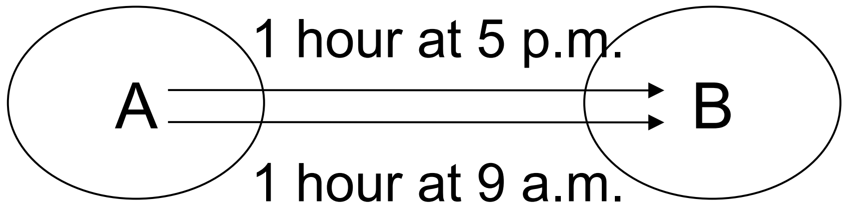
- Routes between two consecutive regions are not relevant

These two movements are not discriminated




- Absolute times are not relevant

These two movements are not discriminated



# Trajectory Clustering

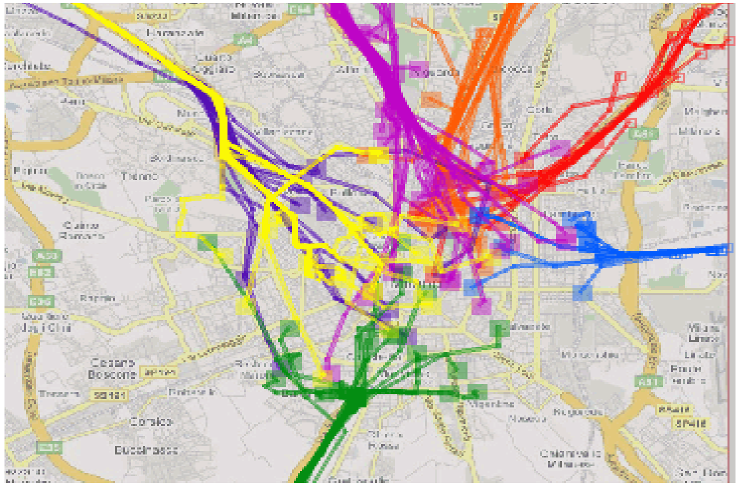
Nanni, Pedreschi.
**Time-focused clustering of trajectories of moving objects**.
J. of Intelligent Information Systems, 2006

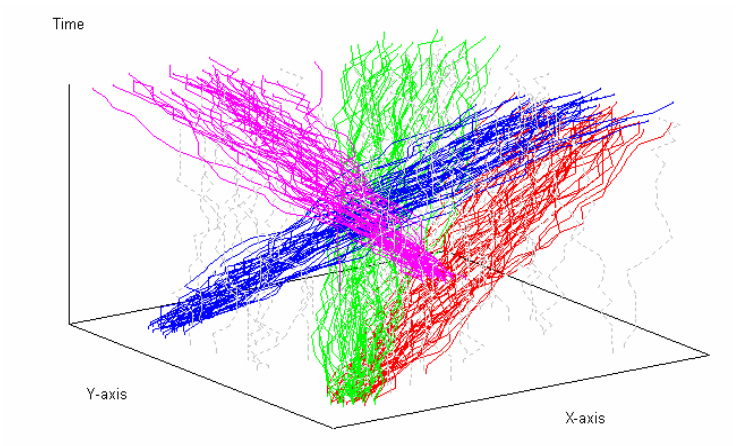
# Clustering: Distance-Based vs. Shape-Based

- Distance-based clustering: Find a group of objects moving together
  - For whole time span
    - high-dimensional clustering
    - probabilistic clustering
  - For partial continuous time span
    - density-based clustering
    - moving cluster, flock, convoy *(borderline case between clustering and patterns)*
  - For partial discrete time span
    - swarm *(borderline case between clustering and patterns)*
- Shape-based clustering: Find similar shape trajectories
  - Variants of shape: translation, rotation, scaling, and transformation
  - Sub-trajectory clustering

# T-clustering

- Trajectories are grouped based on similarity
- Several possible notions of similarity
  - Start/End points
  - Shape of trajectory
  - Shape & time
  - Etc.





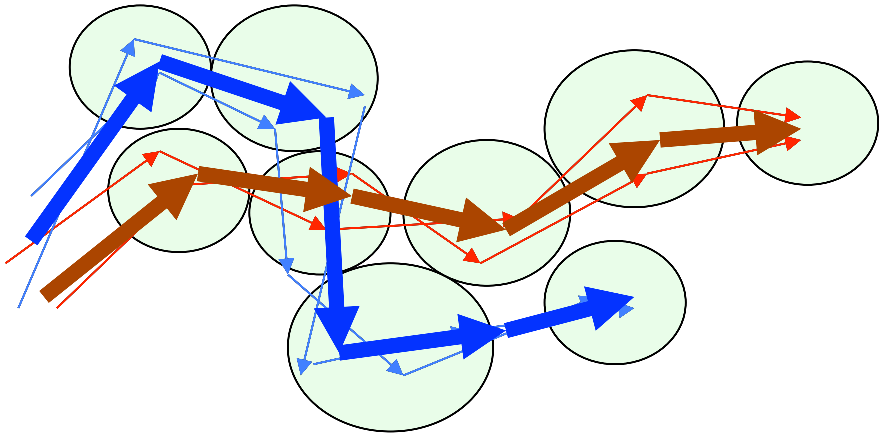

Nanni, Pedreschi. **Time-focused clustering of trajectories of moving objects**. J. of Intelligent Information Systems, 2006.

Rinzivillo, Pedreschi, Nanni, Giannotti, Andrienko, Andrienko. **Visually-driven analysis of movement data by progressive clustering**. J. of Information Visualization, 2008

# Trajectory Clustering



- Questions:
  - Which distance between trajectories?
  - Which kind of clustering?
  - What is a cluster 'mean' in our case?
    - A representative trajectory?



# Which distance?



- Average Euclidean distance (Spatio-temporal distance)

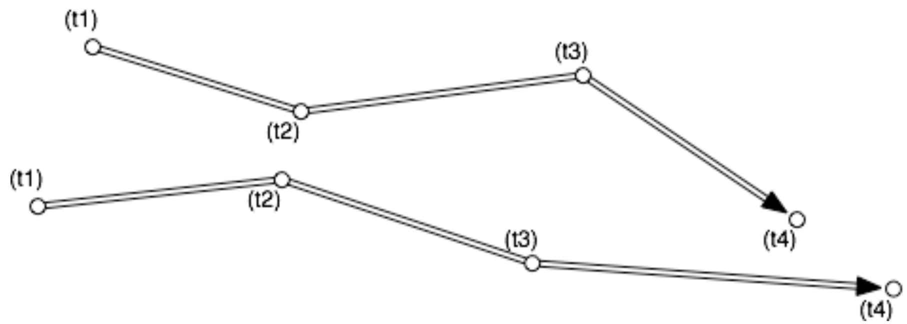
$$D(\tau_1,\tau_2)|_T = \frac{\int_T \boxed{d(\tau_1(t),\tau_2(t))}\, dt}{|T|}$$

distance between moving objects $\tau$1 and $\tau$2 at time *t*

- "Synchronized" behaviour distance
  - Similar objects = almost always in the same place at the same time
- Computed on the whole trajectory
- Computational aspects:
  - Cost = O( |$\tau$1| + |$\tau$2| ) (|$\tau$| = number of points in $\tau$)
  - It is a metric => efficient indexing methods allowed, e.g. [Frentzos et al. 2007]

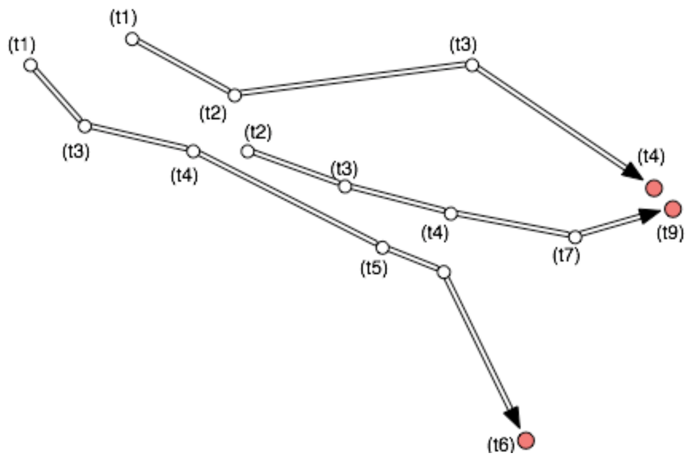
# Average Euclidean Distance Sincronized

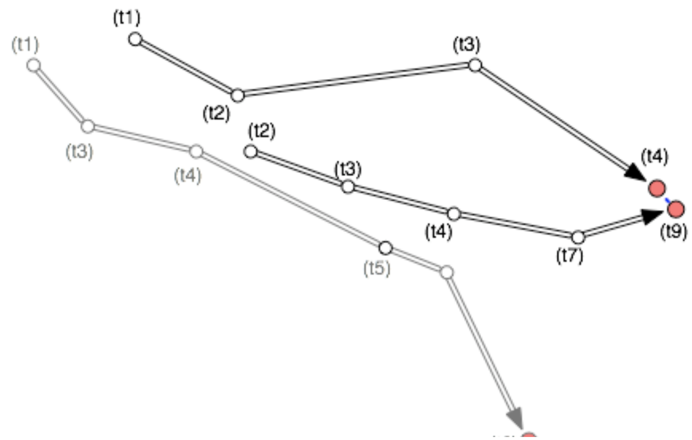
- Align point temporally
- $$D(\tau_1, \tau_2)|_T = \frac{\int_T d(\tau_1(t), \tau_2(t))dt}{|T|}$$
- Eventually assign penalties to non matching points

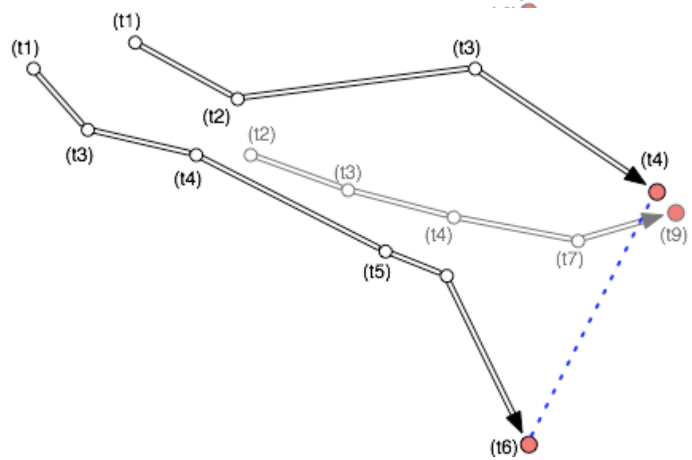


# Common Destination

- Select last point Plast for each trajectory
- D(T,T') = Euclidean(Plast, P'last)

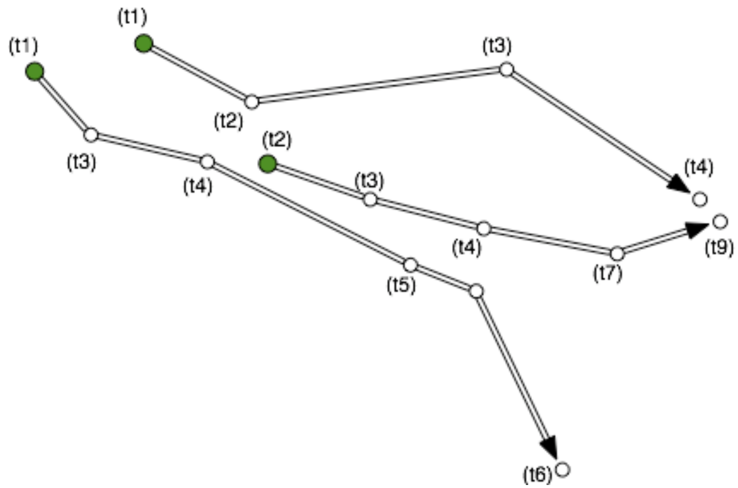


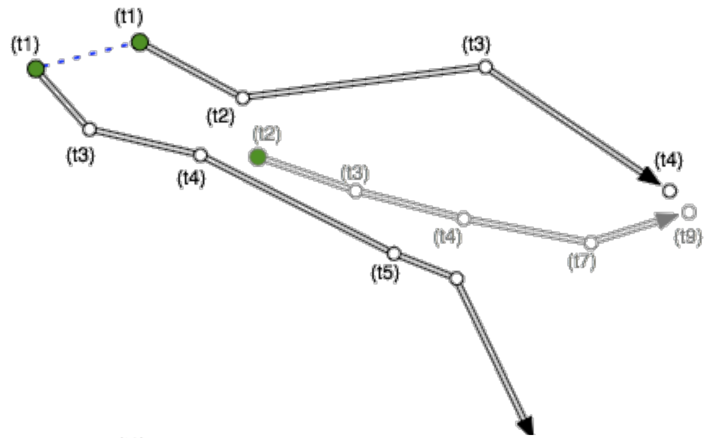



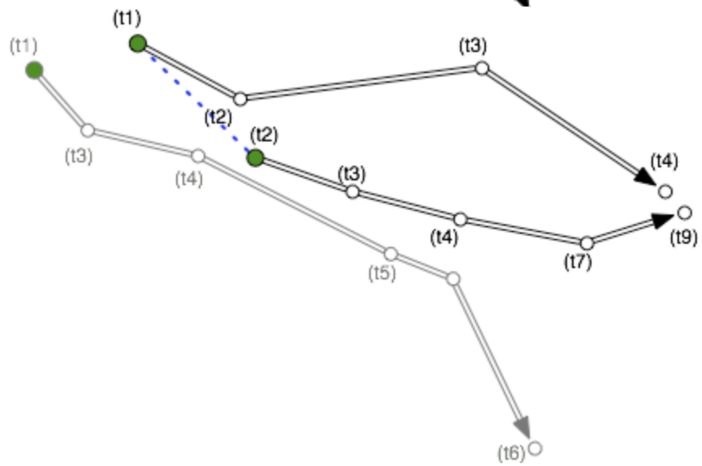



# Common Origins

- Select first point Pfirst for each trajectory
- D(T,T’) = Euclidean(Pfirst, P’first)

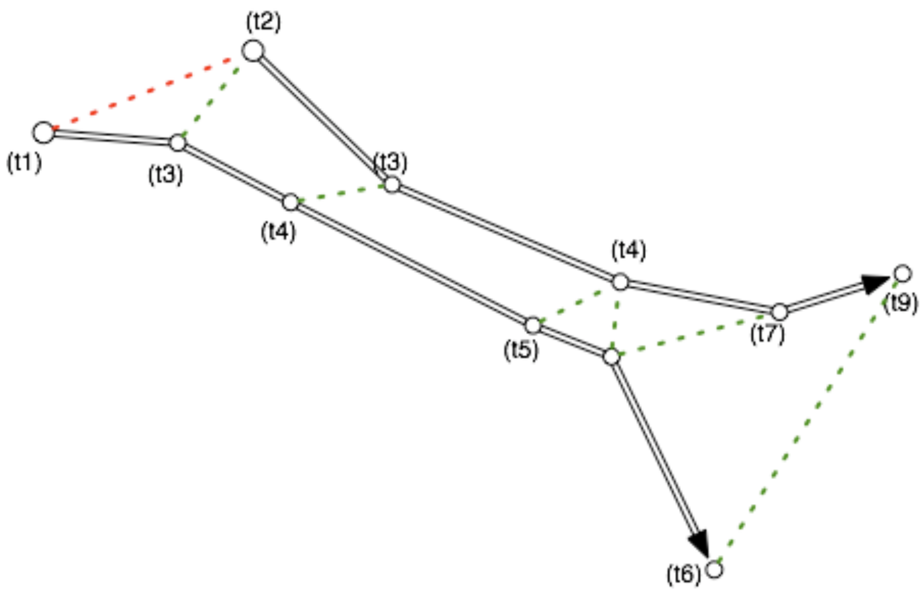


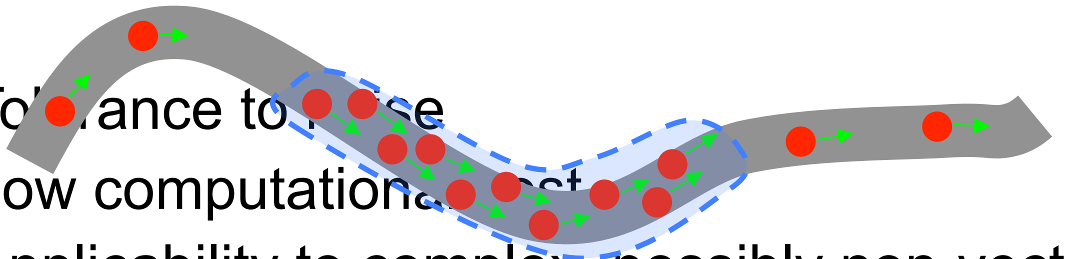
# Route Similarity

- Alignment of points, multiple matches
- Average Euclidean Distance
- Penalties for non matching initial points (no penalties for destinations)



# Which kind of clustering?

- General requirements:
  - Non-spherical clusters should be allowed
    - E.g.: A traffic jam along a road = “snake-shaped” cluster

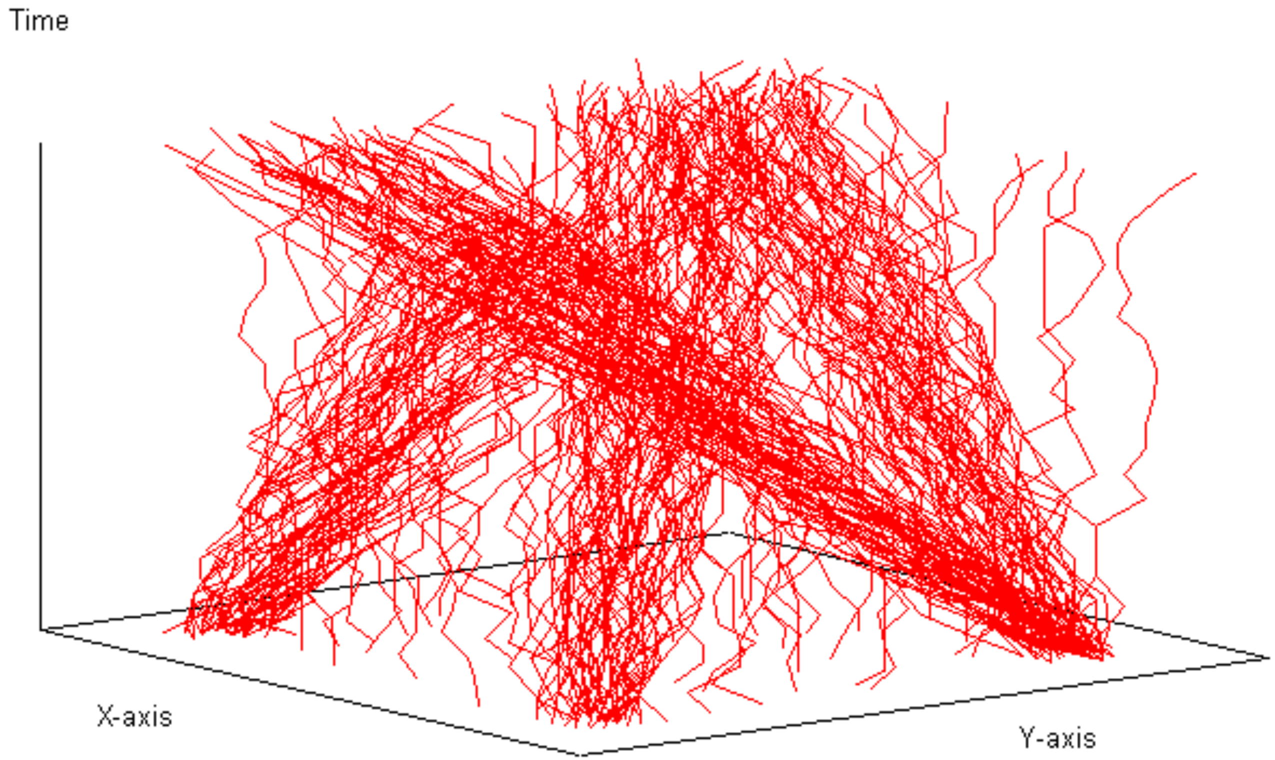


  - Tolerance to noise
  - Low computational cost
  - Applicability to complex, possibly non-vectorial data
- A suitable candidate: Density-based clustering
  - OPTICS (Ankerst et al., 1999)
    - ➔ **T**(rajectory)**-OPTICS**

# A sample dataset



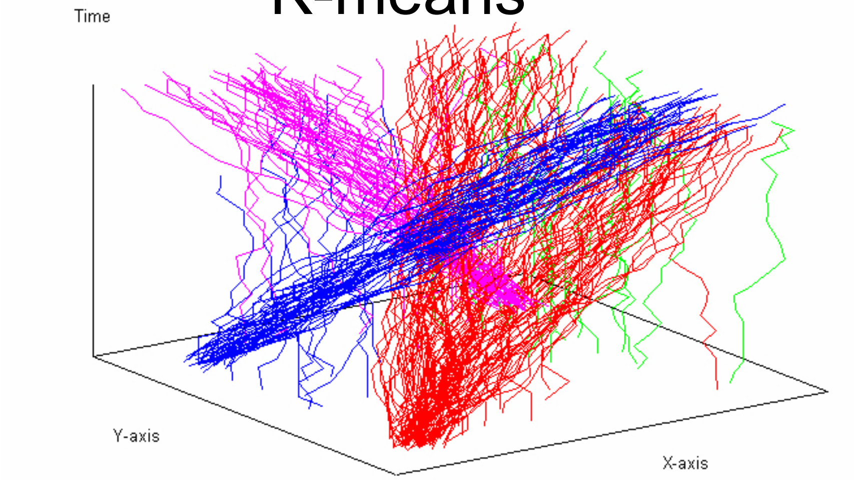
- A set of trajectories forming 4 clusters + noise (syI



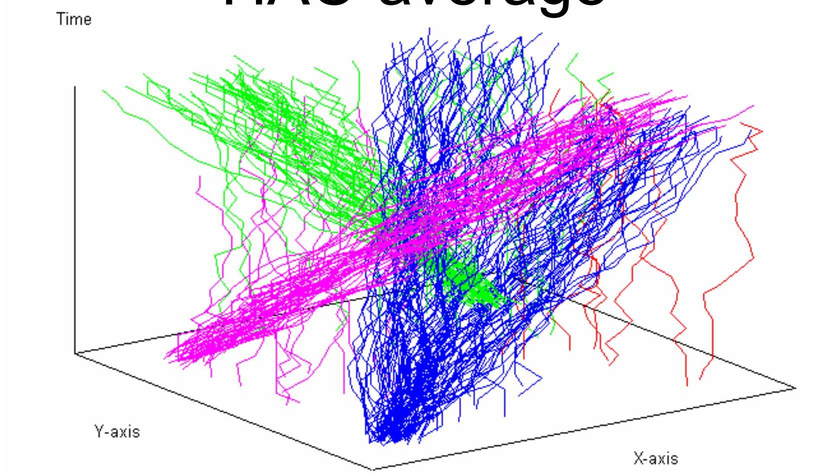
# T-OPTICS vs. HAC & K-means



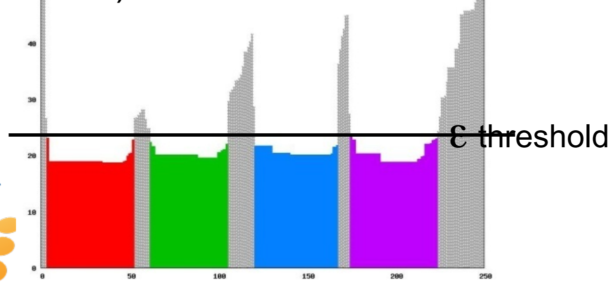
K-means



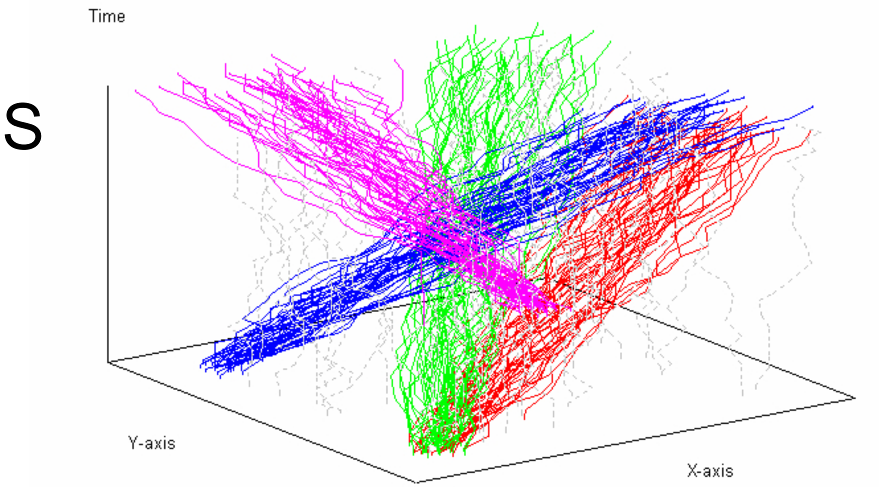

HAC-average




Reachability plot
(= objects reordering for distance distribution)

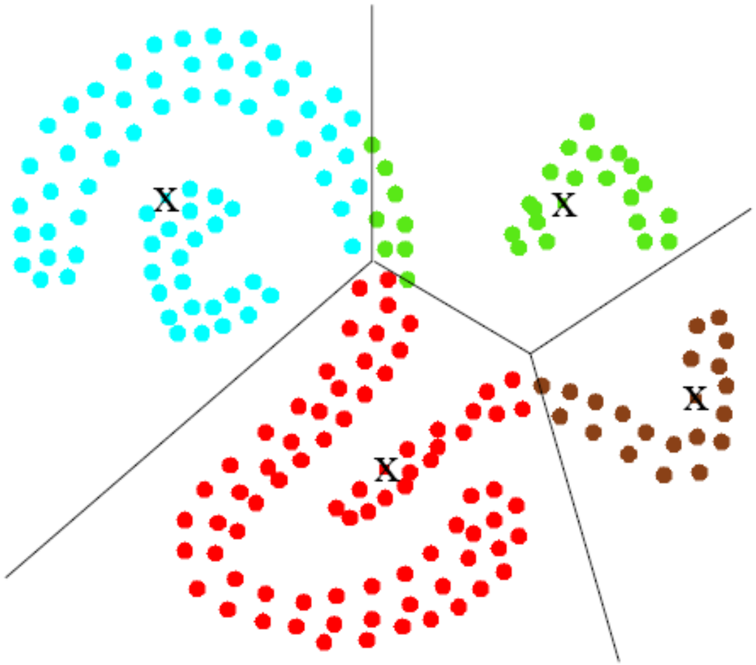


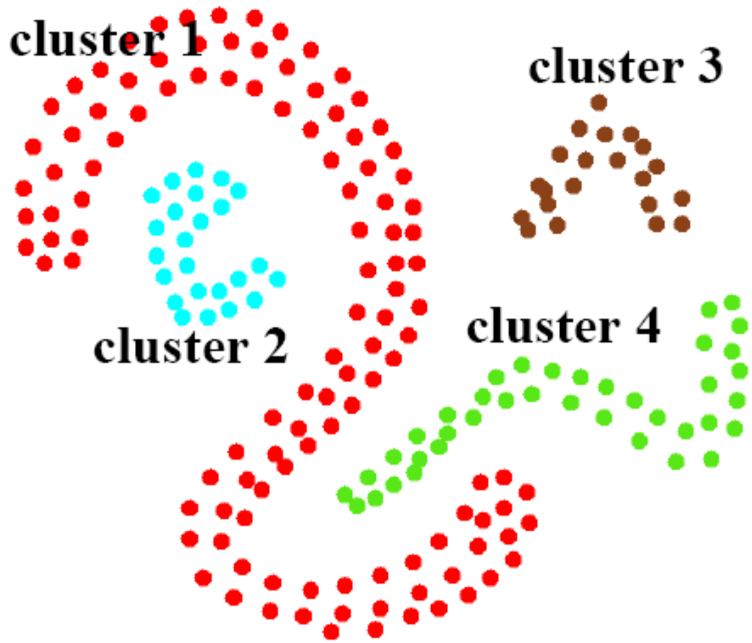

T-OPTICS



# Density Based Clustering

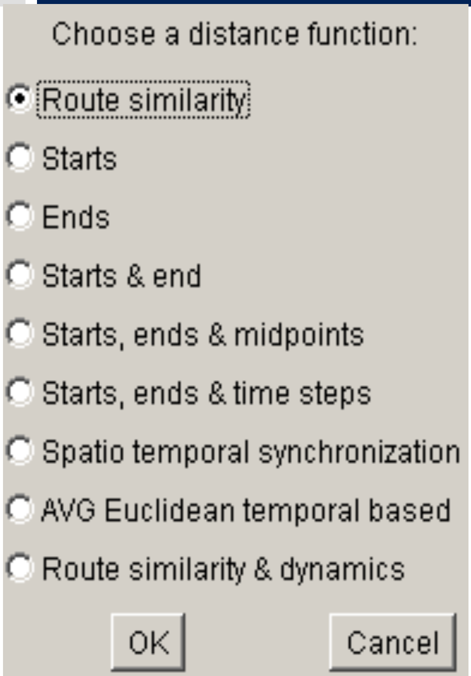
| K-means | Density-based |
| --- | --- |

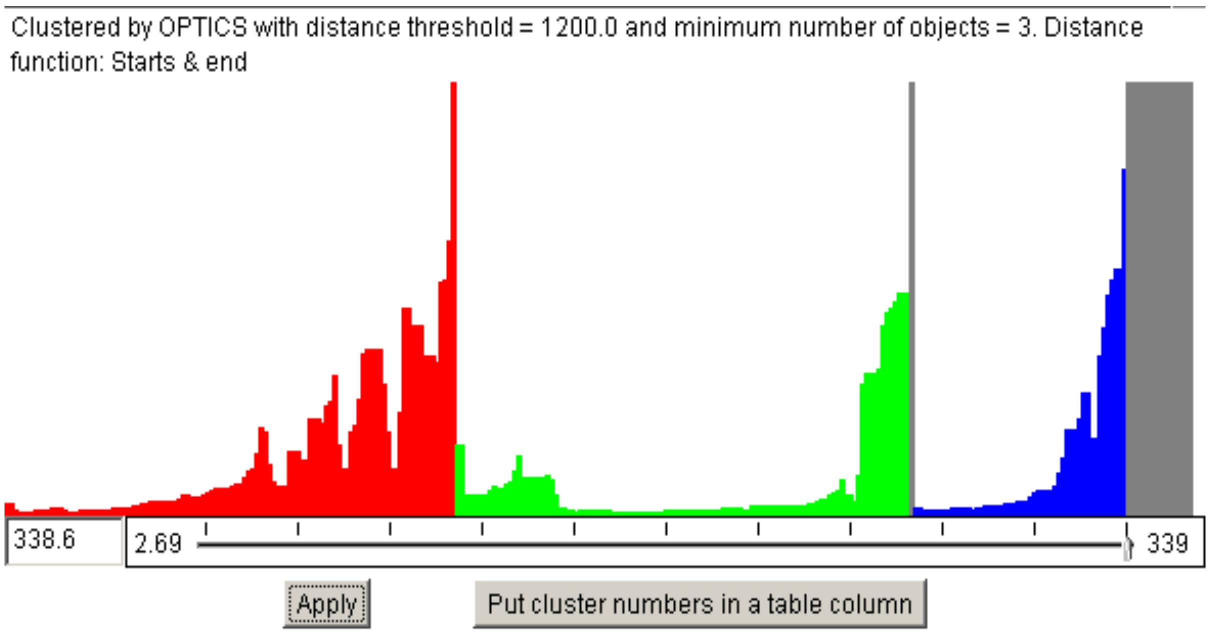


# Interactive density-based trajectory clustering






-

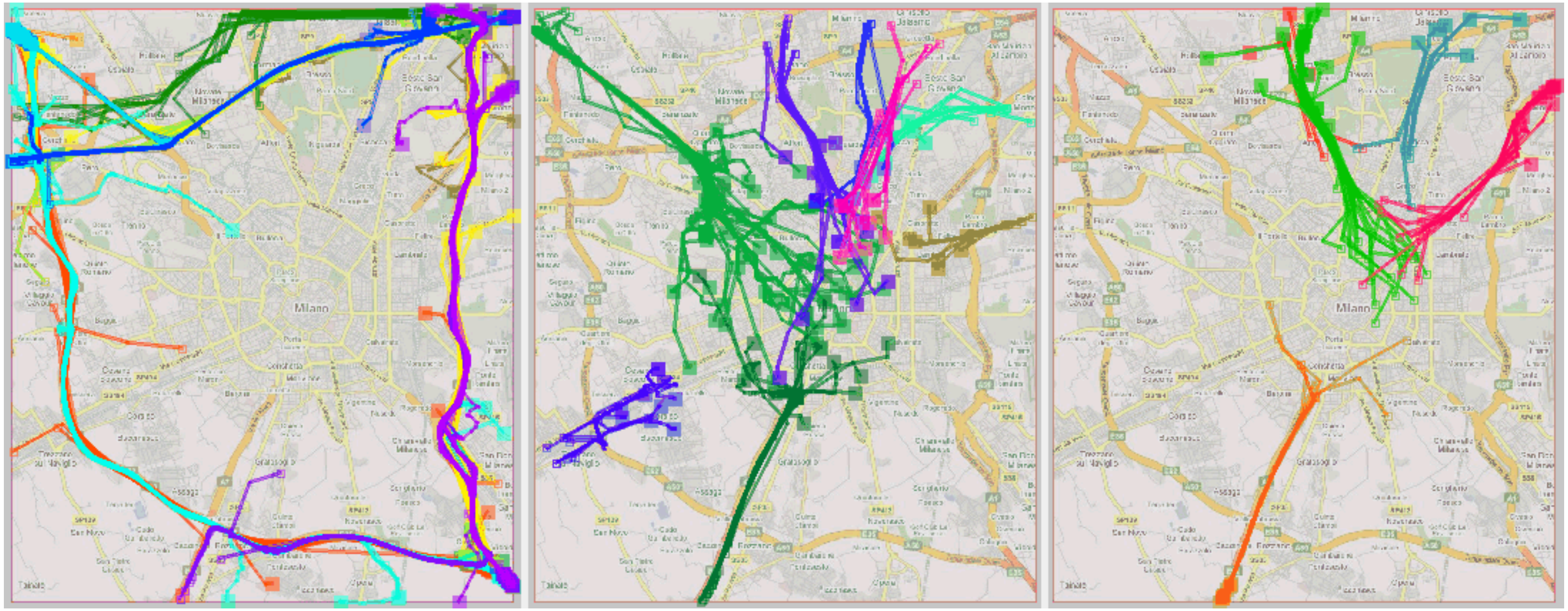
# Ad-hoc distance functions

- Colocation –
  - Link prediction,
  - Semantic behaviors,
  - GSM data
- Spatio-temporal Colocation –
  - Link prediction,
  - Semantic behaviors,
  - GSM data
- Start and End inclusion
  - Car Pooling Matching
- Align to end –
  - Incoming flows
- Align to start –
  - Outcoming flows

# Progressive clustering

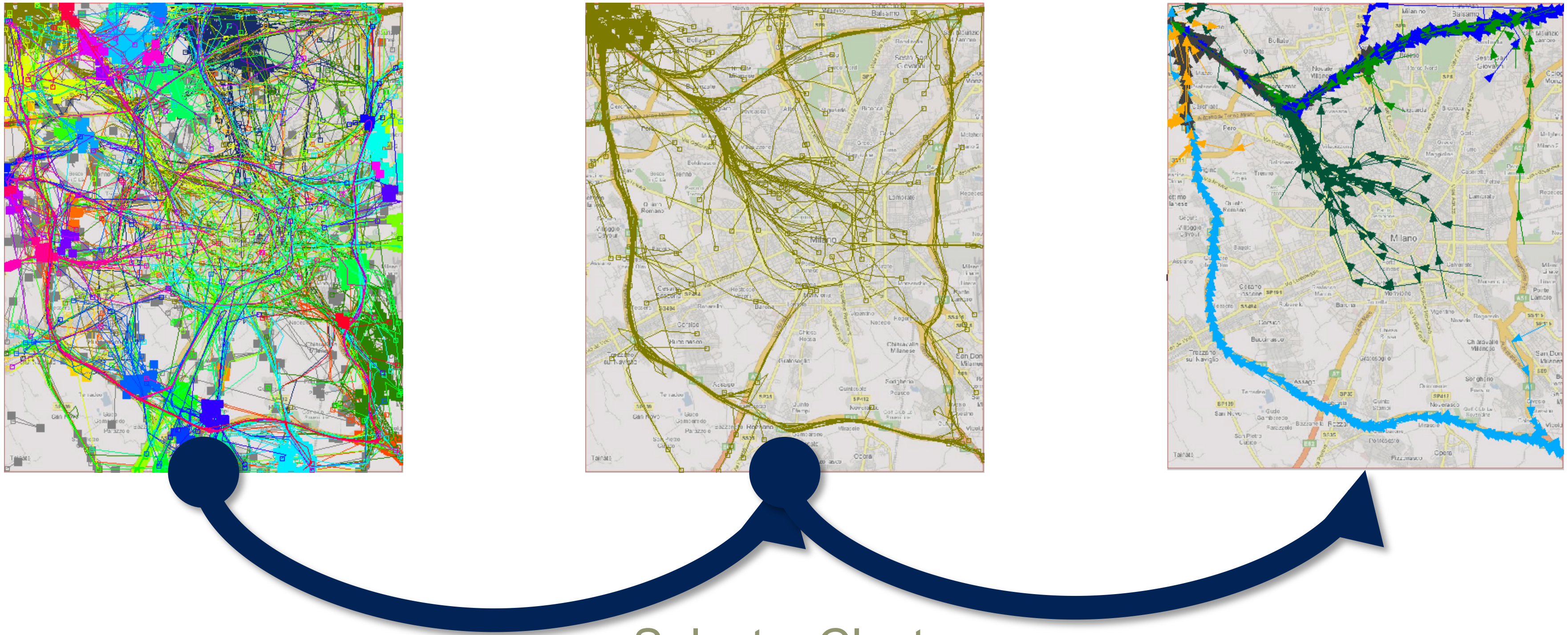


- First, create a large clusters of trajectories using the “common ends” distance function,
- Concentrate on the (big) cluster of inward trajectories (routes towards the city center)
- Refine by creating subclusters using a more sophisticated distance function (route similarity)




Left: peripheral routes; middle: inward routes; right: outward routes.

# Analytical effect of progressive clustering

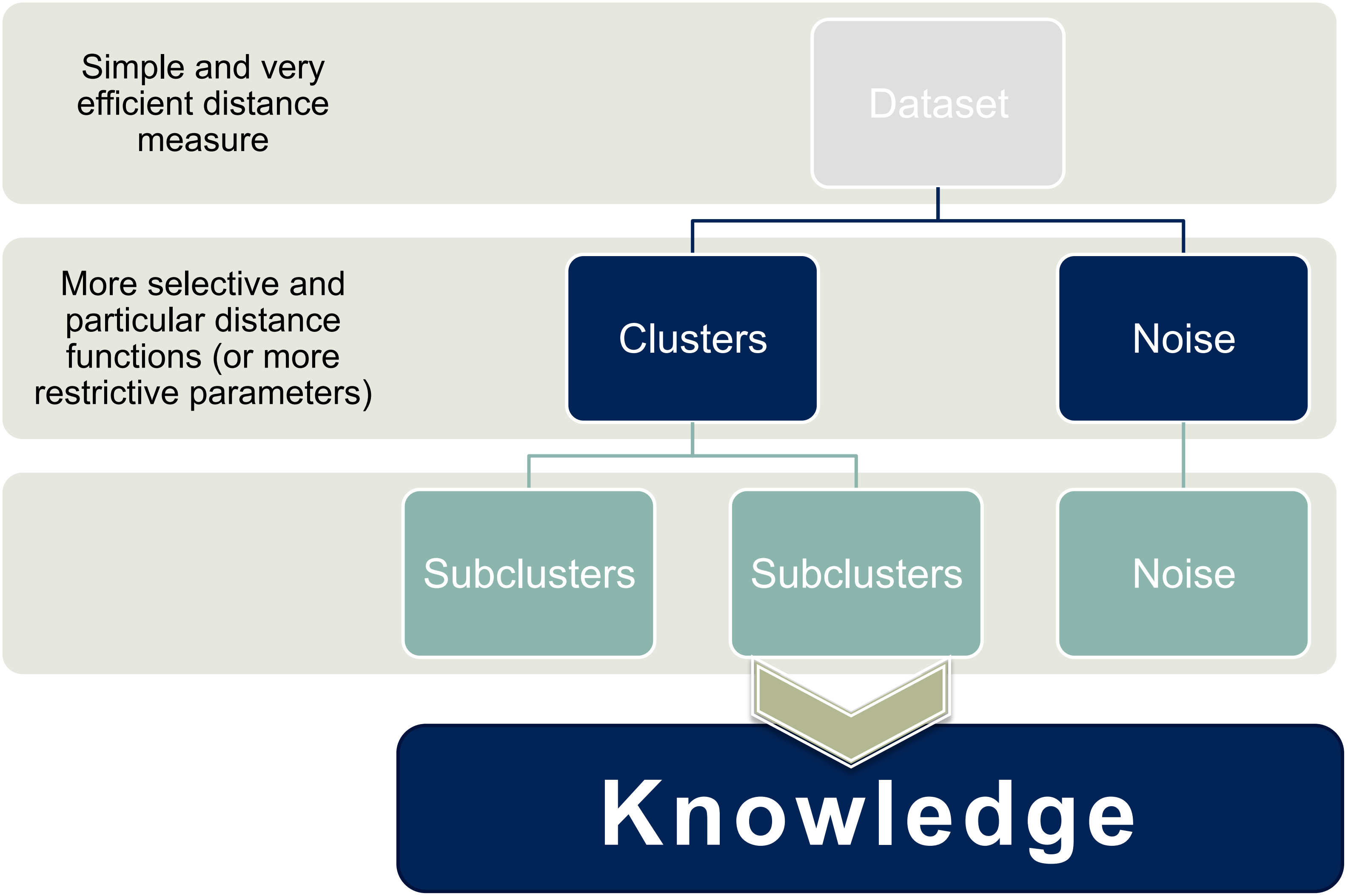


Clustering Data
(Common Destination)

Select a Cluster

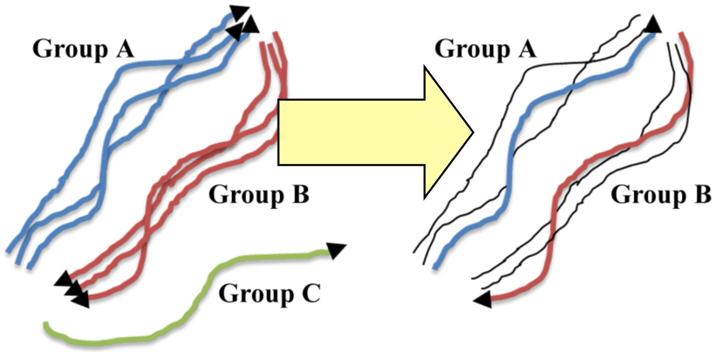
Clustering Data
(route similarity)

# Process Overview

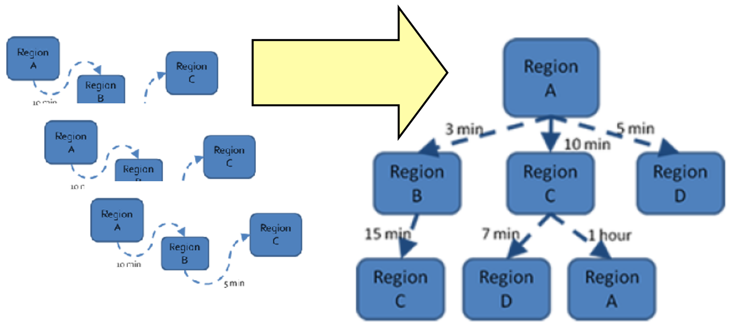


# Derived patterns and models

- Combination & refinement of basic patterns and models




- Individual Mobility Profile: routines consistently followed by a single moving object




- T-PTree: predictive tree built by combining T-Patterns

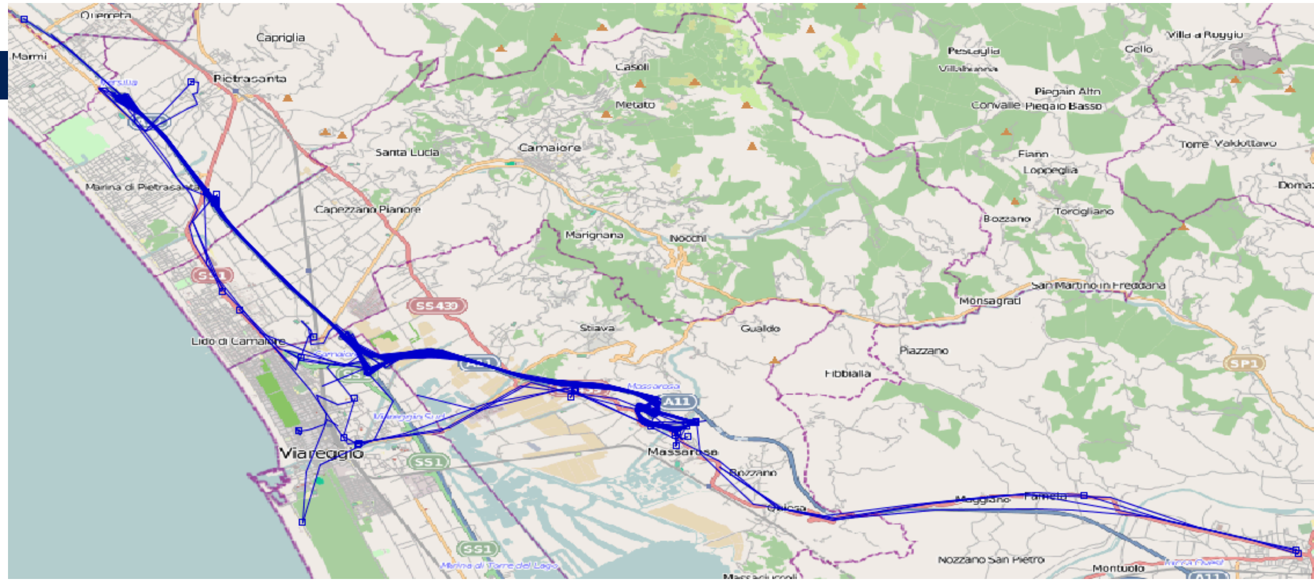
# User's Mobility Profile

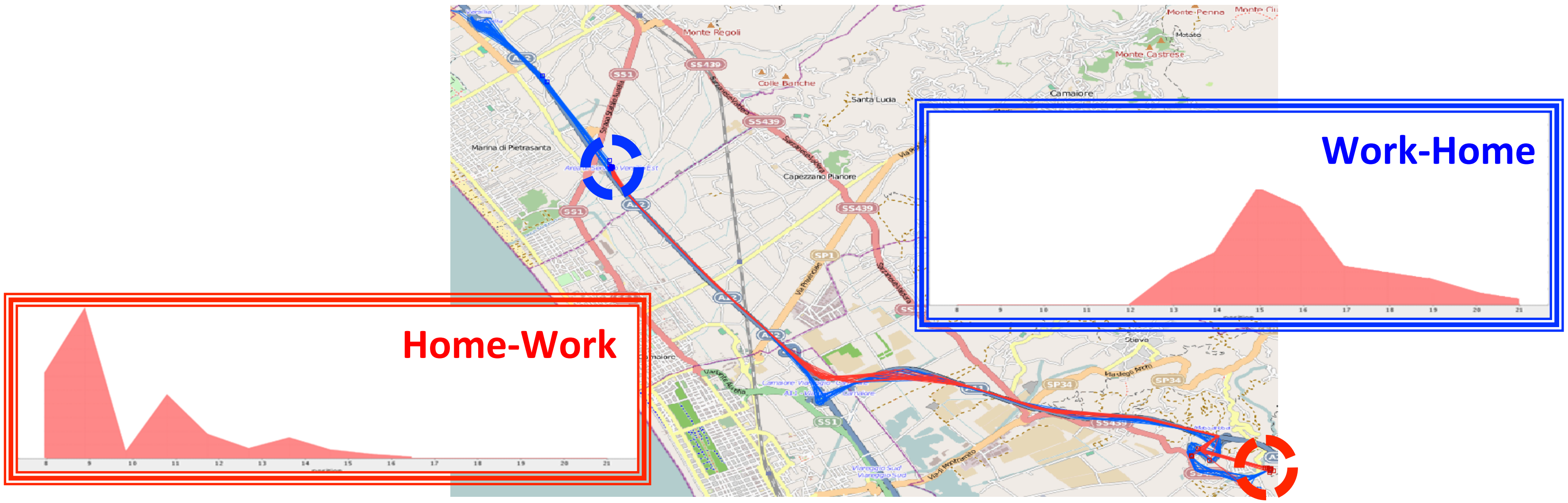
Given the user history as an ordered sequence of spatio-temporal points, we want to extract a set of *routines* in order to create the his\her *mobility profile.*

Where:

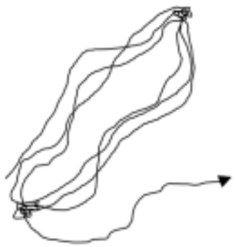
- A *Routine* is a typical local behavior of the user.
- A *Mobility profile* is the set of user's routines

# Discovering individual systematic movements



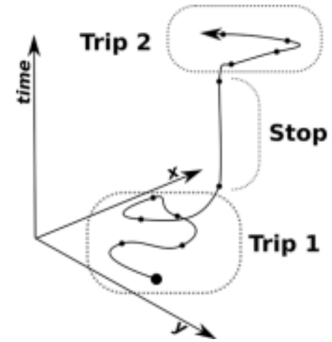



# Derived patterns and models: mobility profiles



## User history

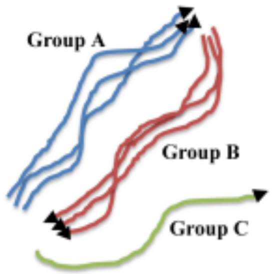
An ordered sequence of spatio-temporal points.




## Trips construction

Cutting the user history when a **stop** is detected

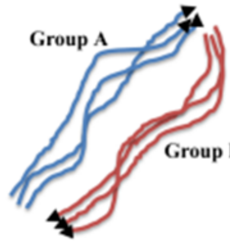
Stops Spatial Threshold
Stops Temporal Threshold




## Grouping

Performing a density based **clustering** equipped with a **spatio temporal** distance function

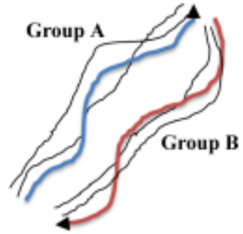
Spatial Tollerance
Temporal Tollerance
Spatio temporal distance




## Pruning

Groups with a small Number of trips are Pruned

Support Threshold




## Profile extraction

The **medoid** of each group becomes user's **routines** and the all set become the user's **mobility profile**

Trasarti, Pinelli, Nanni, Giannotti.
**Mining mobility user profiles for car pooling**. ACM SIGKDD 2011

# Car pooling application

The user profiling might be deployed in a car pooling service which provides a pro-active suggestions without the need for the user to explicitly describe (and update) the trips of interest.

Matching of two routines:

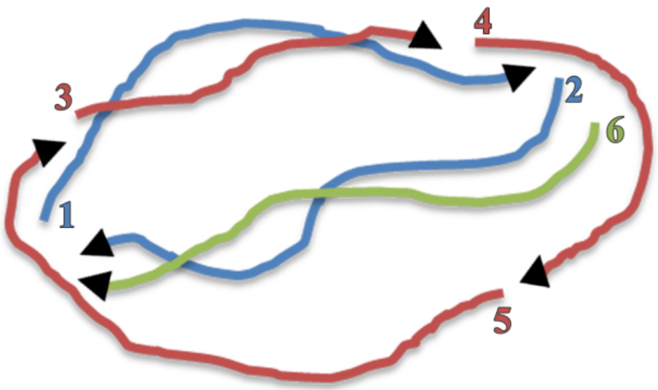
$T_1 = \langle p_1^1 \dots p_n^1 \rangle$ *and* $T_2 = \langle p_1^2 \dots p_m^2 \rangle$

$$
\begin{aligned}
contained(T_1, T_2, th^{walking}_{distance}, th^{wasting}_{time}) \equiv\ & \exists i, j \in \mathcal{N} \mid \\
& 0 < i \leq j \leq m \wedge \\
& Dist(p_1^1, p_i^2) + Dist(p_n^1, p_j^2) \leq th^{walking}_{distance} \wedge \\
& Dur(p_1^1, p_i^2) + Dur(p_n^1, p_j^2) \leq th^{wasting}_{time}
\end{aligned}
$$

Mobility profile share-ability:

*mobility profiles* $\tilde{T}_1$ *and* $\tilde{T}_2$

$$
profileShare(\tilde{T}_1, \tilde{T}_2, th^{walking}_{distance}, th^{wasting}_{time}) = \frac{\left| \left\{ p \in \tilde{T}_1 \mid \exists q \in \tilde{T}_2 . Share(p, q, th^{walking}_{distance}, th^{wasting}_{time}) \right\} \right|}{|\ \tilde{T}_1\ |}
$$

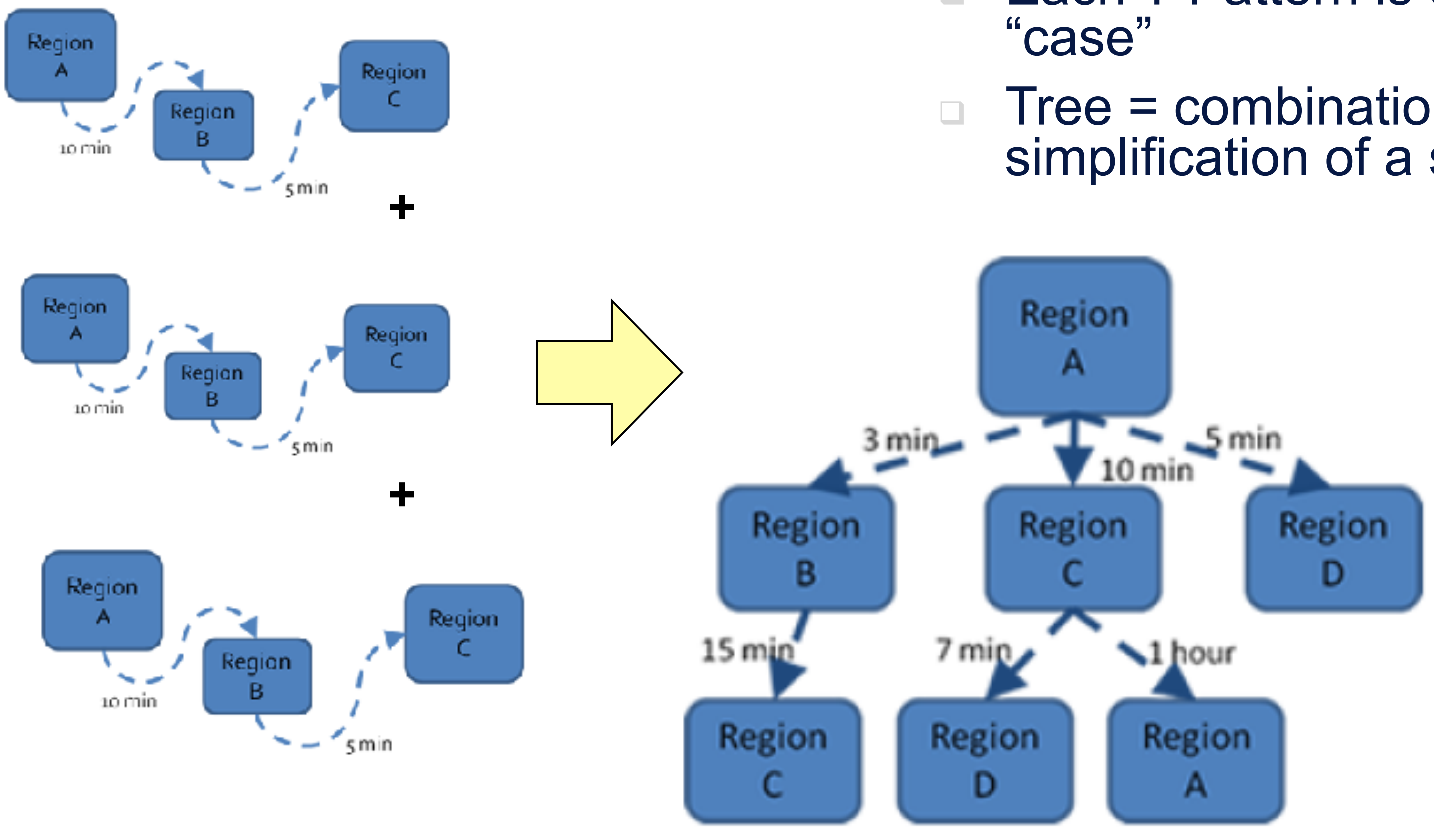



| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | - | - | F | F | F | F |
| 2 | - | - | F | F | F | **T** |
| 3 | **T** | F | - | - | - | F |
| 4 | F | F | - | - | - | F |
| 5 | F | F | - | - | - | F |
| 6 | F | **T** | F | F | F | - |

| | | | |
|---|---|---|---|
| | - | 0 | 1/2 |
| | 1/3 | - | 0 |
| | 1 | 0 | - |

# Derived patterns and models: T-Prediction Tree

- Rule-based prediction model
- Each T-Pattern is used as a "case"
- Tree = combination / simplification of a set of T-

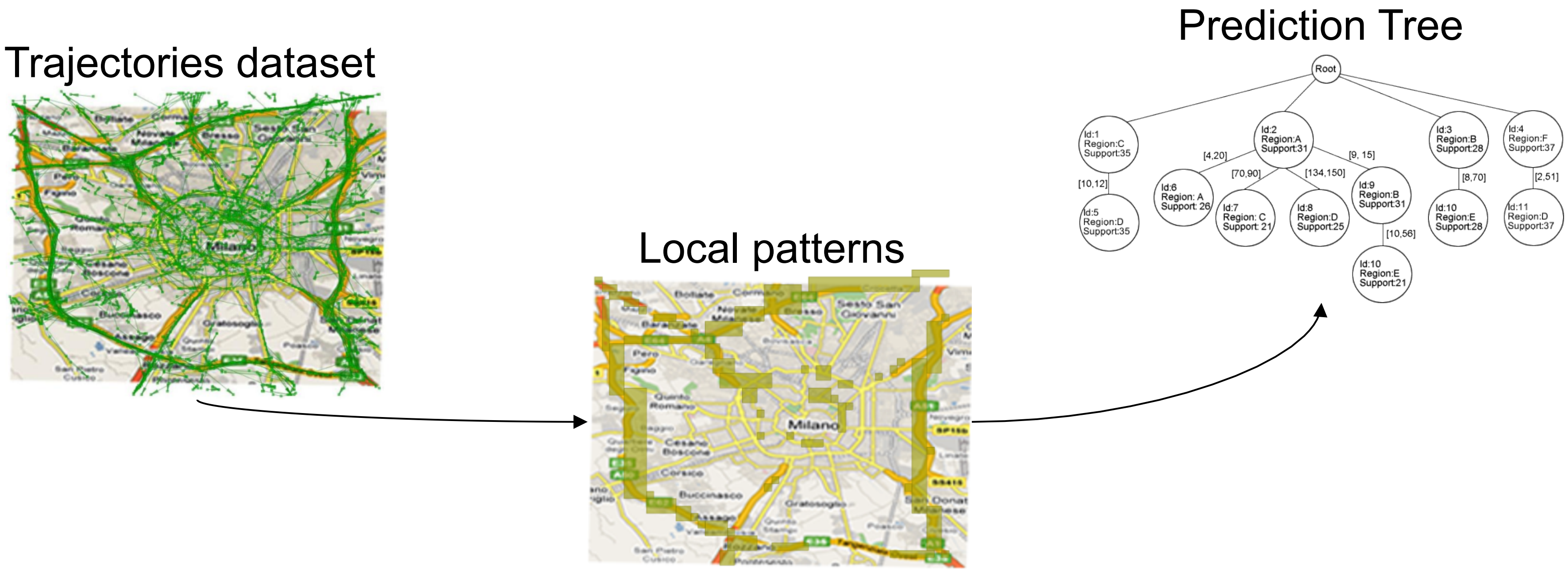



Monreale, Pinelli, Trasarti, Giannotti.
Where Next: a predictor on Trajectory pattern mining. Proc. ACM SIGKDD 2009

# Basic Idea: People move as the crowd moves

How to realize this idea:

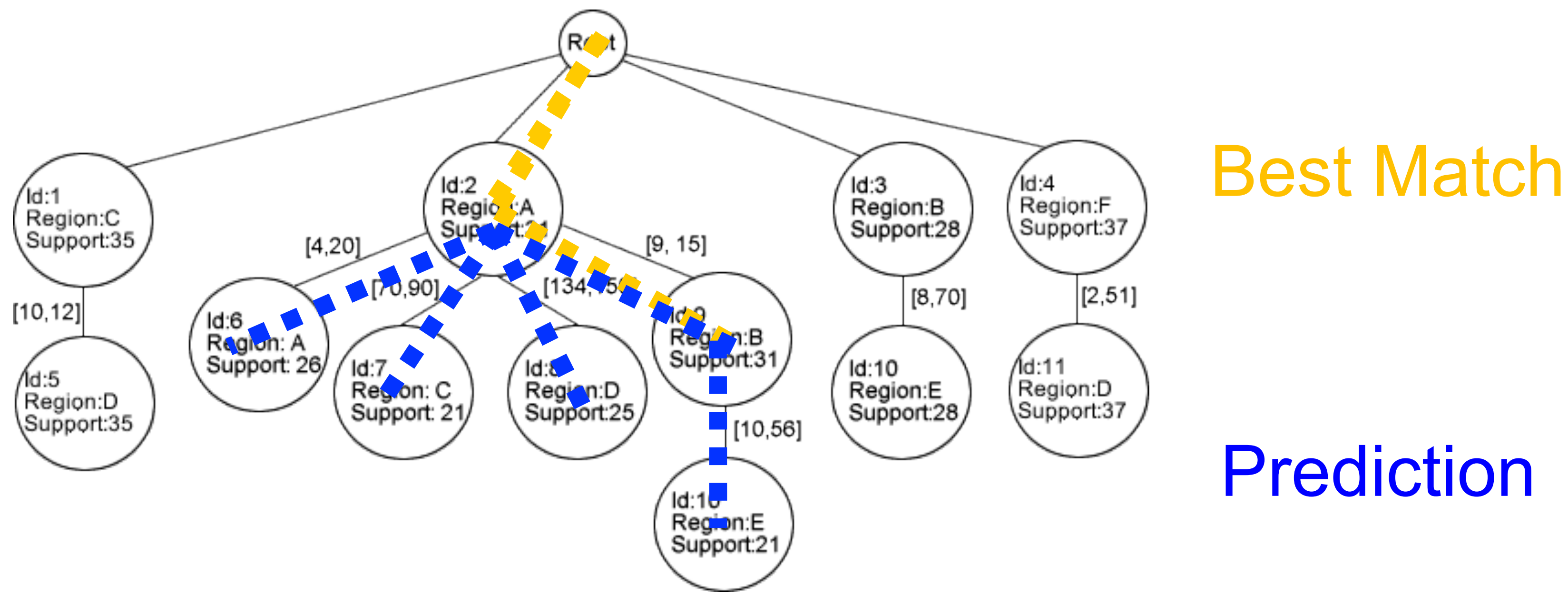
- Extract patterns from **all the available movements** in a certain area instead of on the individual history of an object;
- Using these **Local movement patterns** as predictive rules.
- Build a prediction tree as global model.



# Predict by means of T-Pattern tree

Given a new trajectory:

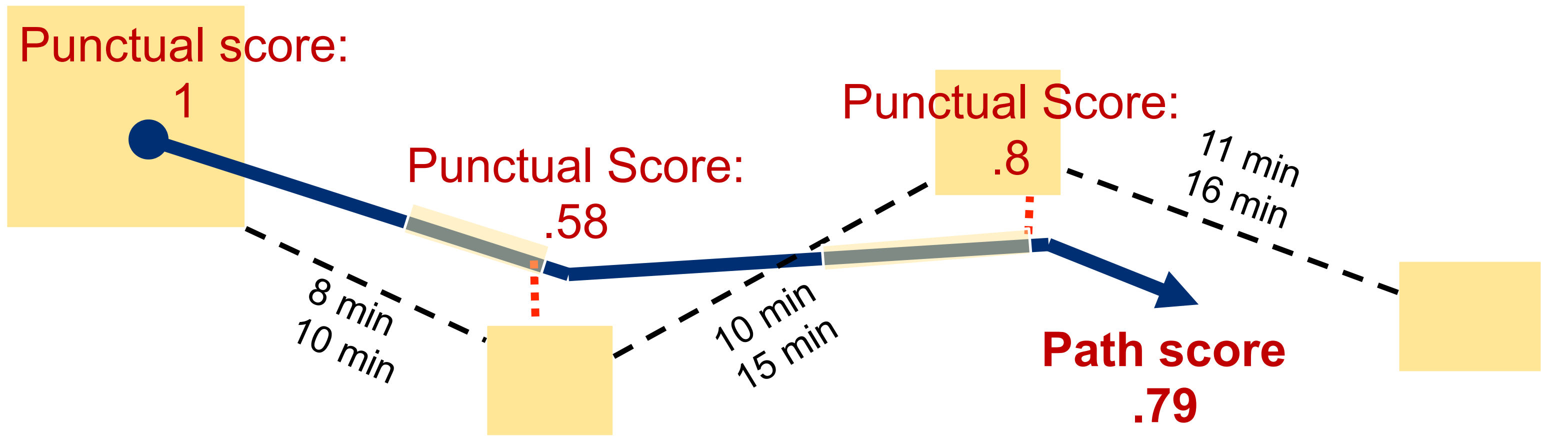
1. Search for best match
2. Candidate generation
3. Make predictions




How to compute the Best Match?

# Computing the path score

The path score is the aggregation of all punctual scores along a path.

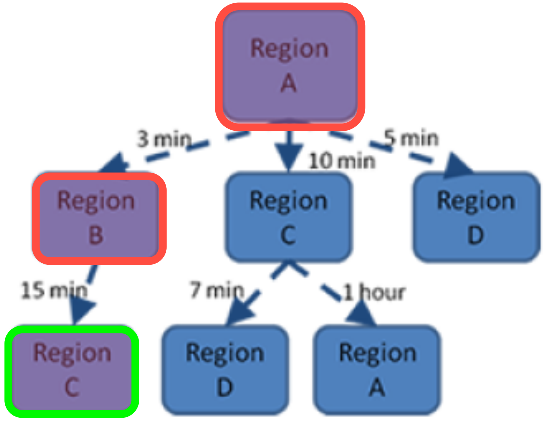


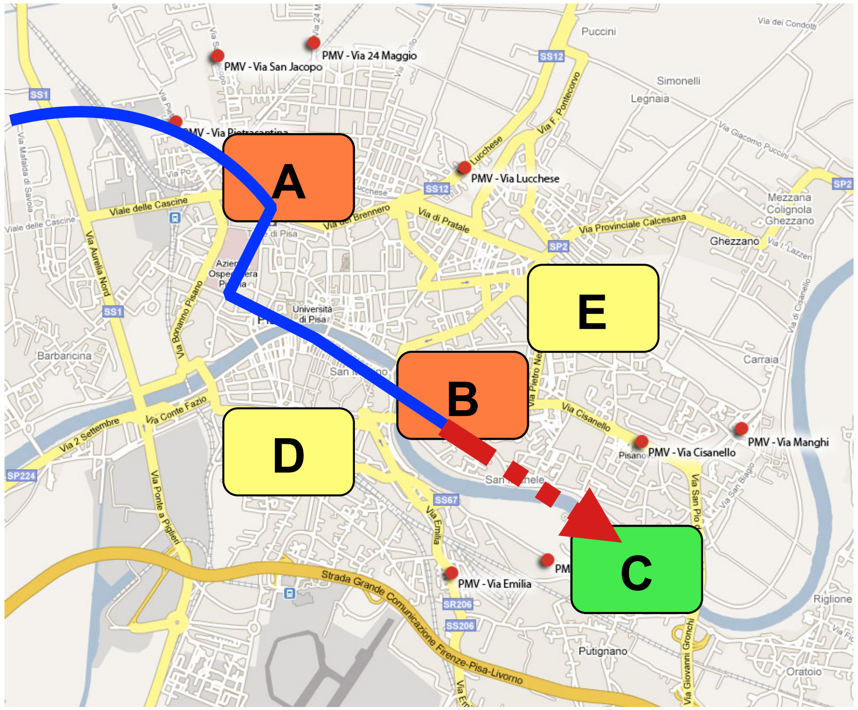

The **Best Match** is the path having:

- ✓ the maximum path score;
- ✓ at least one admissible prediction.

# Derived patterns and models: T-PTree

- Example: Compare actual trajectory against the T-PTree
  - Spatial and temporal similarity used to choose best "rule"

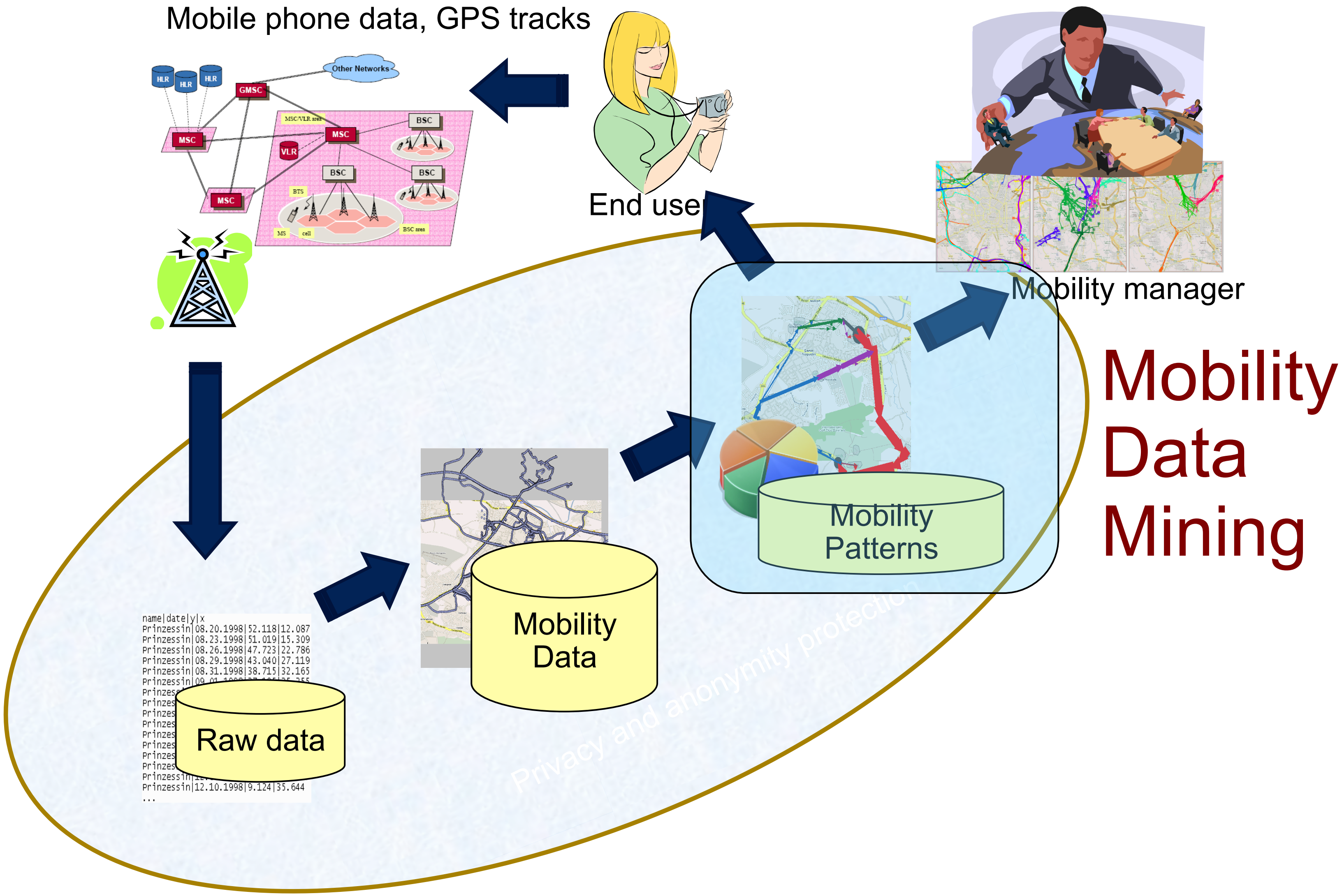


# M-Atlas system

Download from: http://m-atlas.eu

# The (GeoP)KDD process

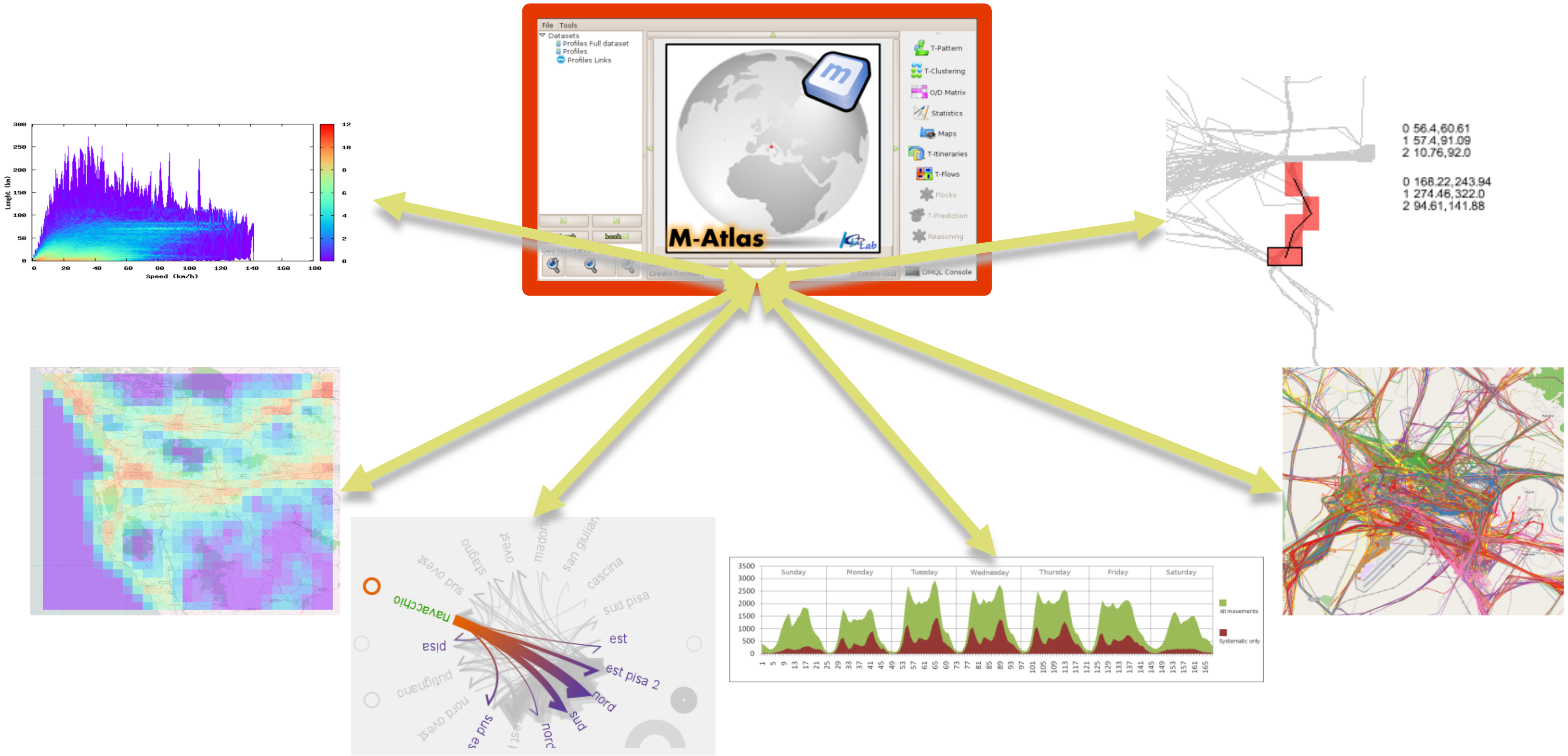


# M-Atlas input

- M-Atlas: An atlas for “urban mobility behaviors”. A framework to query, analyze and navigate the results on mobility data

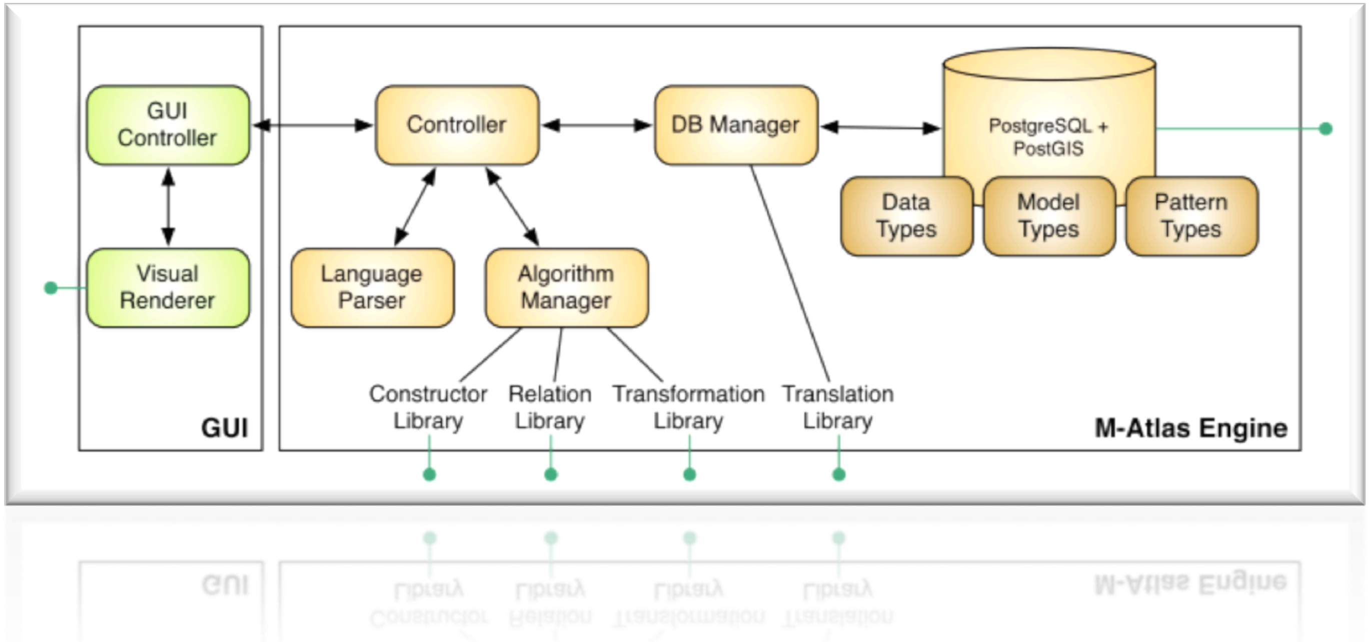
# M-Atlas platform

- A tool kit to extract, store, combine different kinds of models to build mobility knowledge discovery processes.

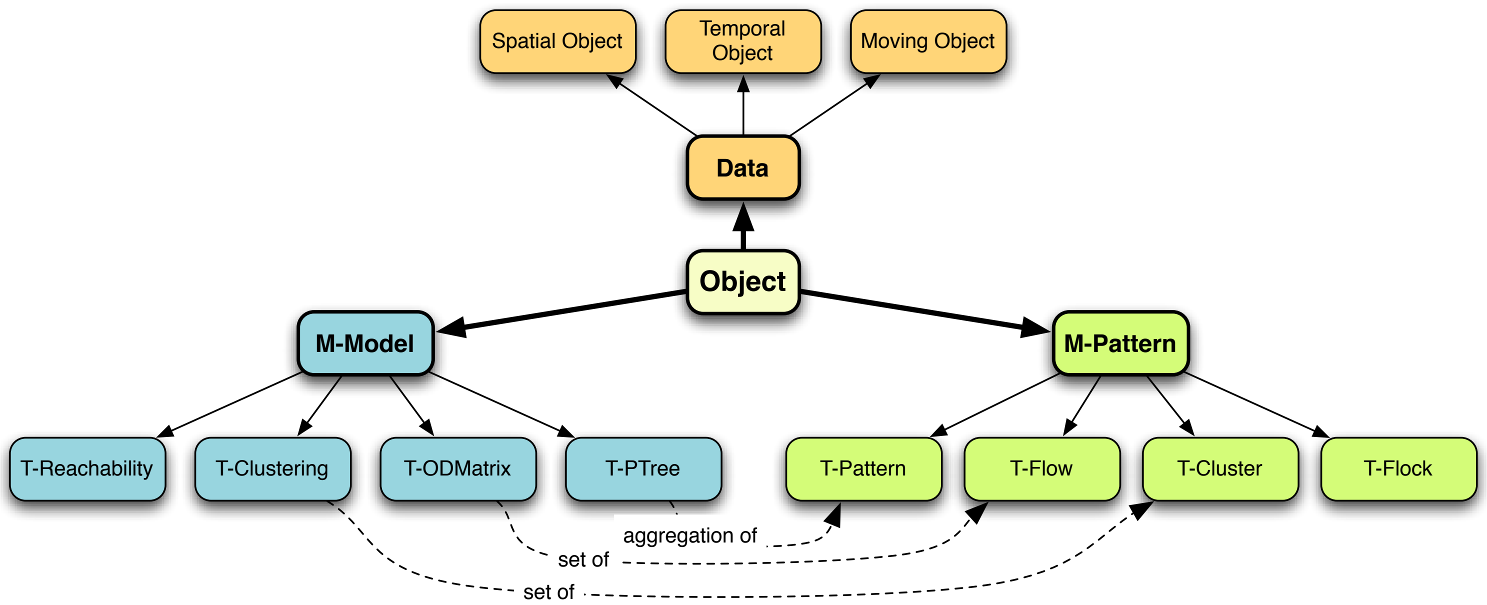


# M-Atlas System

Centralized database which contains all the data, patterns and models. It is possible to extend the system with new algorithms and new data, pattern or model types.



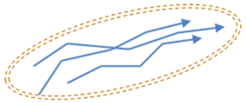
# Objects taxonomy in M-Atlas

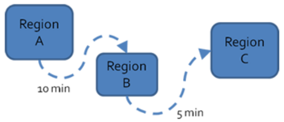


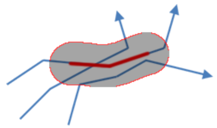

We distinguish between models and patterns: a **pattern** is a representation of a local property that holds over a sub-group of mobility data; a **model** is a representation of a global property that holds over an entire dataset.

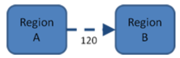
# DMQL: Model constructors

**Patterns**

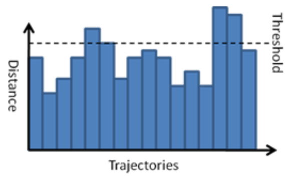


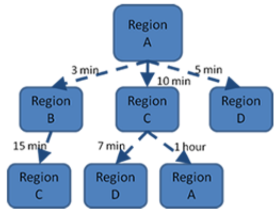


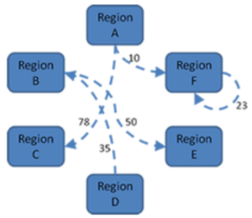



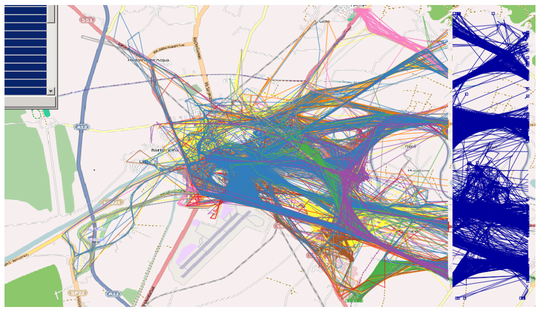



**Models**



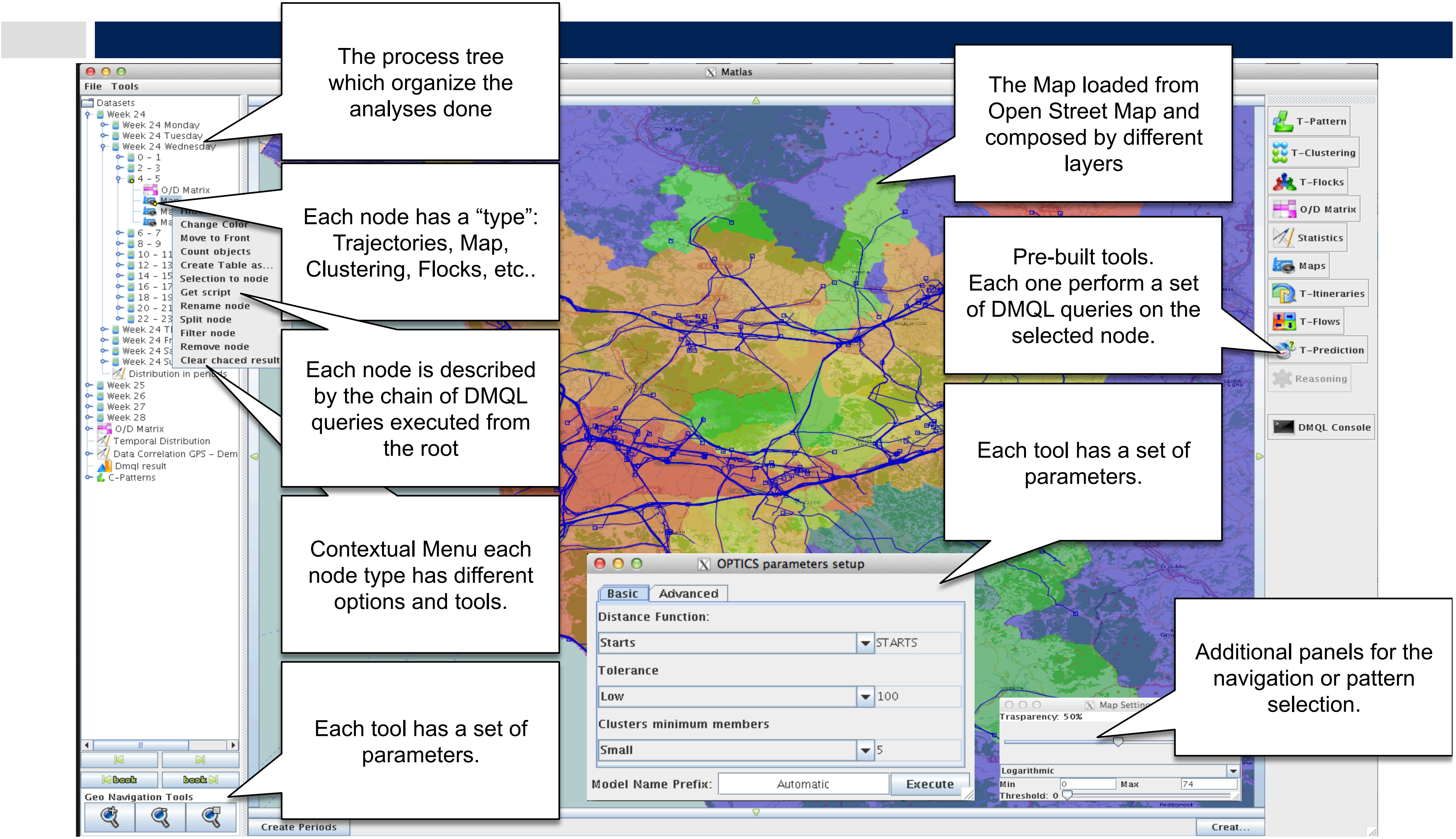



```
CREATE MODEL ClusteringTable MINE AS T-CLUSTERING
FROM (Select t.id, t.trajobj from TrajectoryTable t)
SET T-CLUSTERING.FUNCTION = ROUTE_SIMILARITY AND
    T-CLUSTERING.EPS = 100 AND
    T-CLUSTERING.MIN_PTS = 20
```

# The user Interface



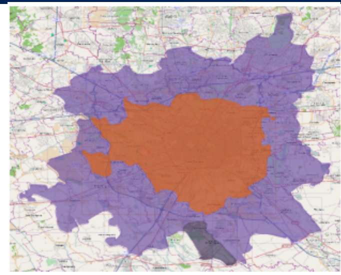
# Mobility Data Mining process as a DMQL query

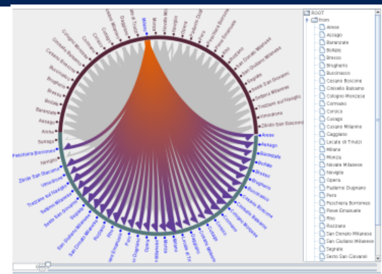
- CREATE MODEL MilanODMatrix AS MINE ODMATRIX
  FROM (SELECT t.id, t.trajectory FROM TrajectoryTable
  t),
  (SELECT orig.id, orig.area FROM MunicipalityTable orig),
  (SELECT dest.id, dest.area FROM MunicipalityTable dest)

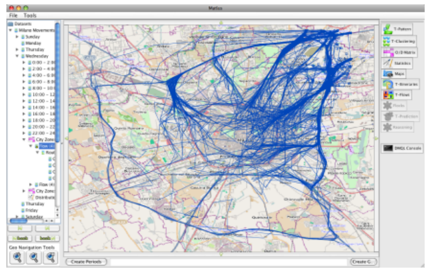
- CREATE RELATION CenterToNESuburbTrajectories
  USING ENTAIL
  FROM (SELECT t.id, t.trajectory FROM TrajectoryTable t,
  MilanODMatrix m
  WHERE m.origin = Milan AND
  m.destination IN (Monza, ..., Brugherio))

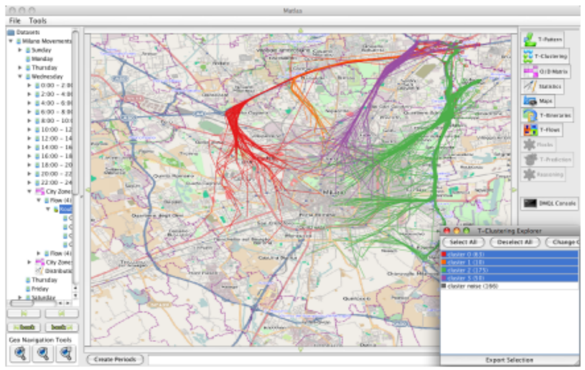
- CREATE MODEL ClusteringTable AS MINE T-
  CLUSTERING
  FROM (Select t.id, t.trajectory from
  CenterToNESuburbTrajectories t)
  SET T-CLUSTERING.FUNCTION = ROUTE_SIMILARITY
  AND
  T-CLUSTERING.EPS = 400 AND
  T-CLUSTERING.MIN_PTS = 5

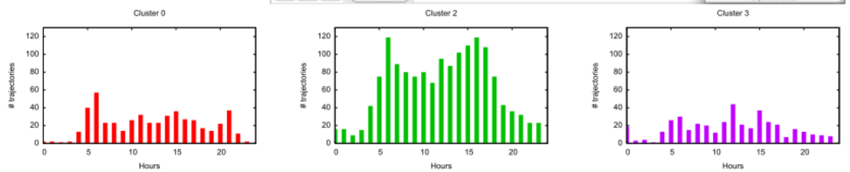
- CREATE RELATION DistributionCluster USING
  CONTAINS
  FROM (SELECT t.id, t.trajectory , c.cid FROM
  ClusteringTable c, TrajectoryTable t WHERE c.tid=t.id),
  (SELECT * FROM Periods p)
  WHERE cid IN (0,2,3)



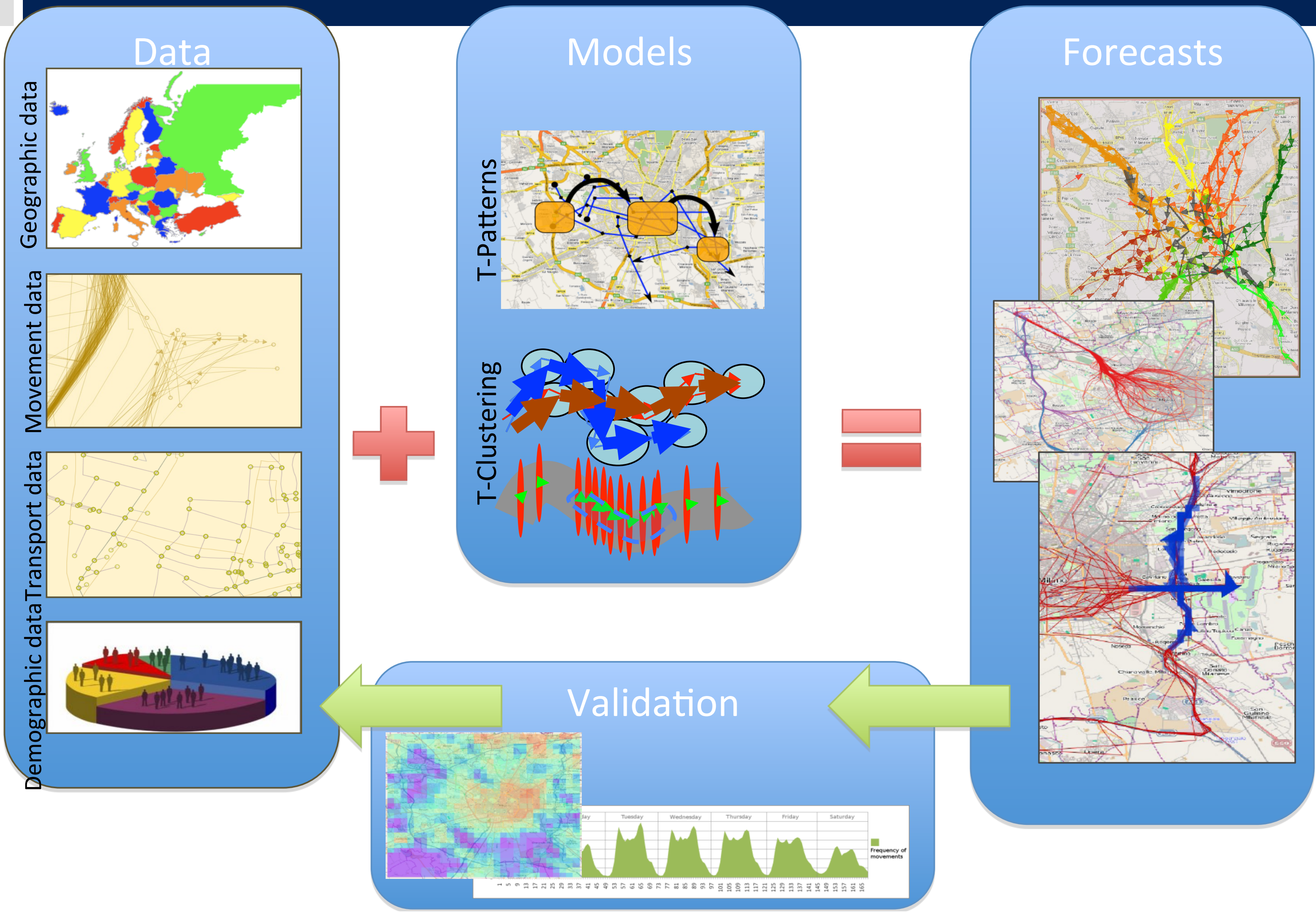








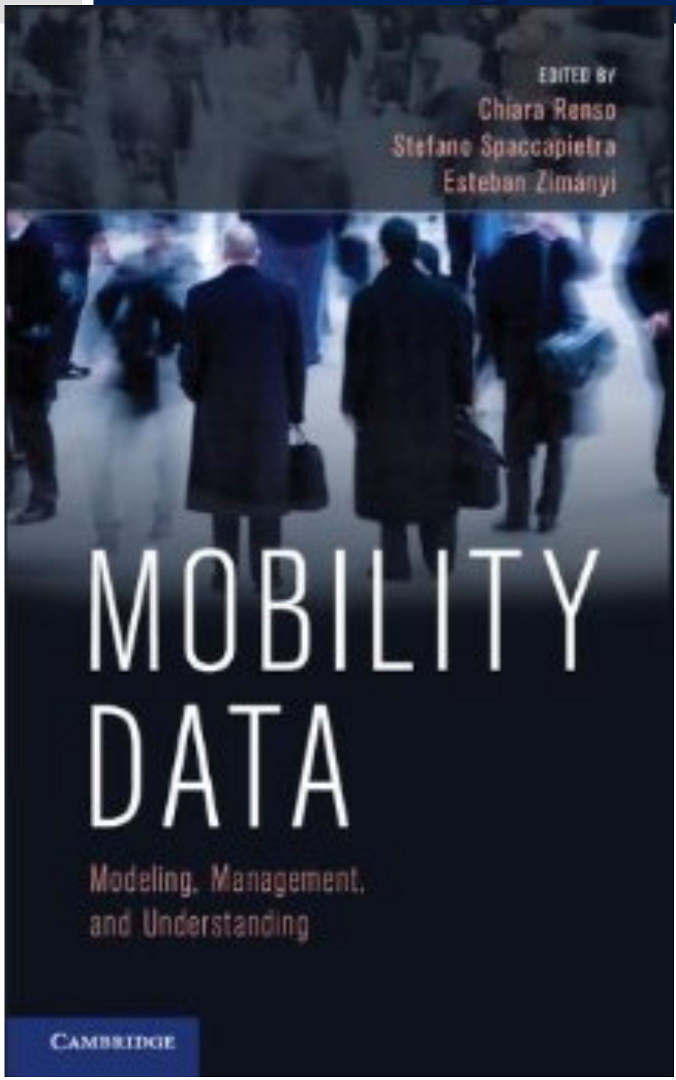
# From DATA to KNOWLEDGE



# Biblio


- Octotelematics s.p.a.: http://www.octotelematics.it/.
- G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti. Interactive visual clustering of large collections of trajectories. In Visual Analytics Science and Technology, 2009. VAST2009. IEEE Symposium on , pages 3{10, 2009.
- F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo, and R. Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data. VLDB J. , 20(5):695{719, 2011.
- Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. Trajectory pattern mining. In KDD , 2007.
- M.Nanni, R.Trasarti, G.Rossetti, and D.Pedreschi. Ecient distributed computation of human mobility aggregates through user mobility proles. In UrbComp12 , 2012.
- Roberto Trasarti, Fabio Pinelli, Mirco Nanni, and Fosca Giannotti. Mining mobility user proles for car pooling. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining , pages 1190{1198. ACM, 2011.
- Monreale, A., Pinelli, F., Trasarti, R., & Giannotti, F. (2009). WhereNext: a location predictor on trajectory pattern mining. In KDD '09 (p. 637-646).
- Mobility Data. Cambridge Press:
- Mobility Data Mining and Privacy

# Mobility data: Modeling, Managing and understanding, Cambridge press.




I. **Mobility Data Modeling and Representation**
- Trajectories and their Representations, S. Spaccapietra, C. Parent, L. Spinsanti
- Trajectory Collection and Reconstruction, G. Marketos, M.L Damiani, N. Pelekis, Y. Theodoridis, Z. Yan
- Trajectory Databases, R.H. Guting, T. Behr, C. Duntgen
- Trajectory Data Warehouses, A.A. Vaisman, E. Zimányi
- Mobility and Uncertainty, C. Silvestri, A.A. Vaisman

II. **Mobility Data Understanding**
- Mobility Data Mining, M. Nanni
- Understanding Human Mobility using Mobility Data Mining, C. Renso, R. Trasarti
- Visual Analytics of Movement: A Rich Palette of Techniques to Enable Understanding, N. Andrienko, G. Andrienko
- Mobility Data and Privacy, F. Giannotti, A. Monreale, D. Pedreschi

**III. Mobility Applications**
- Car Traffic Monitoring, D. Janssens, M. Nanni, S. Rinzivillo
- Maritime Monitoring, T. Devogele, L. Etienne, C. Ray
- Air Traffic Analysis, C. Hurter, G. Andrienko, N. Andrienko, R.H. Guting, M Sakr
- Animal Movement, S. Focardi, F. Cagnacci
- Person Monitoring with Bluetooth Tracking, M. Versichele, T. Neutens, N. Van de Weghe

**IV. Future Challenges and Conclusions**
- A Complexity Science Perspective on Human Mobility, F. Giannotti, L. Pappalardo, D. Pedreschi, D. Wang
- Mobility and Geo-Social Networks, L. Spinsanti, M. Berlingerio, L. Pappalardo

**Conclusions**, C. Renso, S. Spaccapietra, E. Zimànyi

Giannotti · Pedreschi (Eds.)

## Mobility, Data Mining and Privacy

The technologies of mobile communications and ubiquitous computing pervade our society, and wireless networks sense the movement of people and vehicles, generating large volumes of mobility data. This is a scenario of great opportunities and risks: on one side, mining this data can produce useful knowledge, supporting sustainable mobility and intelligent transportation systems; on the other side, individual privacy is at risk, as the mobility data contain sensitive personal information. A new multidisciplinary research area is emerging at this crossroads of mobility, data mining, and privacy.

This book assesses this research frontier from a computer science perspective, investigating the various scientific and technological issues, open problems, and roadmap. The editors manage a research project called GeoPKDD, Geographic Privacy-Aware Knowledge Discovery and Delivery, funded by the EU Commission and involving 40 researchers from 7 countries, and this book tightly integrates and relates their findings in 13 chapters covering all related subjects, including the concepts of movement data and knowledge discovery from movement data; privacy-aware geographic knowledge discovery; wireless network and next-generation mobile technologies; trajectory data models, systems and warehouses; privacy and security aspects of technologies and related regulations; querying, mining and reasoning on spatiotemporal data; and visual analytics methods for movement data.

This book will benefit researchers and practitioners in the related areas of computer science, geography, social science, statistics, law, telecommunications and transportation engineering.

ISBN 978-3-540-75176-2

9 783540 751762

› springer.com

Giannotti
Pedreschi (Eds.)

Mobility, Data Mining
and Privacy

Fosca Giannotti
Dino Pedreschi (Eds.)

# Mobility, Data Mining and Privacy

### Geographic Knowledge Discovery

Springer

# References: Moving Object Databases and Queries


- R. H. Gueting and M. Schneider. *Moving Objects Databases*. Morgan Kaufmann, 2005.
- S. R. Jeffrey, G. Alonso, M.J. Franklin, W. Hong, and J. Widom. A pipelined framework for online cleaning of sensor data streams. *ICDE'06.*
- N. Jing, Y.-W. Huang, and E. A. Rundensteiner. Hierarchical optimization of optimal path finding for transportation applications. *CIKM'96*.
- C. S. Jensen, D. Lin, and B. C. Ooi. Query and update efficient b+-tree based indexing of moving objects. *VLDB'04*.
- E. Kanoulas, Y. Du, T. Xia, and D. Zhang. Finding fastest paths on a road network with speed patterns. *ICDE'06*.
- J. Krumm and E. Horvitz. Predestination: Inferring destinations from partial trajectories. *Ubicomp'06.*
- E. M. Knorr, R. T. Ng, and V. Tucakov. Distance-based outliers: Algorithms and applications. *The VLDB Journal*, 8(3):237-253, February 2000.
- L. Liao, D. Fox, and H. Kautz. Learning and inferring transportation routines. *AAAI'04*.

# References on Moving Object Pattern Mining (I)


- M. Andersson, Gudmundsson, J., Laube, P. & Wolle, T., "Reporting Leaders and Followers Among Trajectories of Moving Point Objects" , GeoInformatica, 2008.
- M. Benkert, J. Gudmundsson, F. Hubner, and T. Wolle. Reporting flock patterns. *Euro. Symp. Algorithms'06.*
- M. Benkert, J. Gudmundsson, F. Hubner, and T. Wolle. Reporting leadership patterns among trajectories. *SAC'07*.
- H. Cao, N. Mamoulis, and D. W. Cheung. Mining frequent spatiotemporal sequential patterns. *ICDM'05*.
- M. G. Elfeky, W. G. Aref, and A. K. Elmagarmid. Periodicity detection in time series databases. IEEE Trans. Knowl. Data Eng., 17(7), 2005.
- R. Fraile and S. J. Maybank. Vehicle trajectory approximation and classification. In *Proc. British Machine Vision Conf.*, Southampton, UK, September 1998.
- F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi. Trajectory pattern mining. KDD'07.
- S. Gaffney and P. Smyth. Trajectory clustering with mixtures of regression models. *KDD'99*.
- J. Gudmundsson and M. J. van Kreveld. Computing longest duration flocks in trajectory data. GIS'06.
- J. Gudmundsson, M. J. van Kreveld, and B. Speckmann. Efficient detection of patterns in 2d trajectories of moving points. *GeoInformatica*, 11(2):195-215, June 2007.
- H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen and H. T. Shen, "Discovery of Convoys in Trajectory Databases", VLDB 2008.

# References on Moving Object Pattern Mining (II)


- P. Kalnis, N. Mamoulis, and S. Bakiras. On discovering moving clusters in spatiotemporal data. SSTD 2005.
- V. Kostov, J. Ozawa, M. Yoshioka, and T. Kudoh. Travel destination prediction using frequent crossing pattern from driving history. *Proc. Int. IEEE Conf. Intelligent Transportation Systems*, Vienna, Austria, Sept. 2005
- Y. Li, J. Han, and J. Yang. Clustering moving objects. *KDD'04*.
- Z. Li, et al., “MoveMine: Mining Moving Object Databases", SIGMOD’10 (system demo)
- Z. Li, B. Ding, J. Han, and R. Kays, “Swarm: Mining Relaxed Temporal Moving Object Clusters”, in submission
- Z. Li, B. Ding, J. Han, and R. Kays, “Mining Hidden Periodic Behaviors for Moving Objects”, in submission
- N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. W. Cheung. Mining, indexing, and querying historical spatiotemporal data. KDD 2004..
- M. Nanni and D. Pedreschi. Time-focused clustering of trajectories of moving objects. *JIIS*, 27 (3), 2006.
- I. F. Sbalzariniy, J. Theriot, and P. Koumoutsakos. Machine learning for biological trajectory classification applications. In *Proc. 2002 Summer Program, Center for Turbulence Research*, August 2002.
- I. Tsoukatos and D. Gunopulos. Efficient mining of spatiotemporal patterns. *SSTD'01*.

# References on Outlier Detection


- E. Horvitz, J. Apacible, R. Sarin, and L. Liao. Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service. *UAI'05*
- E. M. Knorr, R. T. Ng, and V. Tucakov. Distance-based outliers: Algorithms and applications. The VLDB Journal, 8(3):237-253, February 2000.
- J.-G. Lee, J. Han, and X. Li, "Trajectory Outlier Detection: A Partition-and-Detect Framework", ICDE 2008
- J.-G. Lee, J. Han, and K.-Y. Whang, "Trajectory Clustering: A Partition-and-Group Framework", SIGMOD'07
- X. Li, J.  Han, S. Kim, "Motion-alert: Automatic anomaly detection in massive moving objects", ISI 2006
- X. Li, J. Han, S. Kim, and H. Gonzalez, "ROAM: Rule- and Motif-Based Anomaly Detection in Massive Moving Object Data Sets", SDM'07
- J. Owens and A. Hunter. Application of the self-organizing map to trajectory classification. In . 3rd IEEE Int. Workshop on Visual Surveillance, Dublin, Ireland, July 2000